

Erwartungen und Einschätzungen von Lehrkräften zur Leistung von Schüler*innen
Akkuratheit und Bedeutung für Leistungsentwicklung

Dissertation
zur Erlangung des akademischen Grades
doctor philosophiae (Dr. phil.)
im Fach Erziehungswissenschaften

eingereicht am 23. März 2020

an der Kultur-, Sozial- und Bildungswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von **Frau Sarah Gentrup** (Master of Arts)

Prof. Dr.-Ing. Dr. Sabine Kunst
Präsidentin der
Humboldt-Universität zu Berlin

Prof. Dr. Christian Kassung
Dekan der Kultur-, Sozial- und
Bildungswissenschaftlichen Fakultät

Gutachter*innen:

1. Prof. Dr. Hans Anand Pant
2. Prof. Dr. Lysann Zander

Tag der Disputation: 26. Juni 2020

Danksagung

An dieser Stelle möchte ich den vielen Menschen danken, die diese Dissertation ermöglicht und mich auf dem Weg begleitet haben.

Zunächst gilt mein Dank Prof. Dr. Petra Stanat für die fachliche Betreuung der Arbeit sowie das zugleich fordernde und fördernde Umfeld, in dem ich mich während meiner Dissertationsphase bewegen durfte. Durch die vielfältigen Anregungen und Gestaltungsfreiräume, die sie mir eröffnet hat, konnten sich meine Fähigkeiten umfassend weiterentwickeln. Für die Bereitschaft, meine Dissertation zu begutachten, möchte ich Prof. Dr. Hans Anand Pant und Prof. Dr. Lysann Zander meinen herzlichen Dank aussprechen. Ebenfalls bedanken möchte ich mich bei Prof. Dr. Cornelia Kristen und Prof. Dr. Irena Kogan für die anregende Zusammenarbeit im KuL-Projekt sowie an den gemeinsamen Publikationen.

Mein besonderer Dank gilt Dr. Georg Lorenz für den stets bereichernden fachlichen Austausch und die sehr angenehme Zusammenarbeit im KuL-Projekt. Ohne ihn wären die wochenlangen Dienstreisen dröge und zermürend gewesen. Dieser Dank geht auch an meine KuL-Kollegin Susanne Rahmann und an unsere tollen studentischen Mitarbeiterinnen Fenna tom Dieck, Susanne Sebal, Miriam Fendt und Claudia Hencke. Ebenso möchte ich „meiner PostDoc“ Dr. Camilla Rjosk für ihren bedingungslosen Beistand, ihre ermutigenden Worte, hilfreichen Ratschläge sowie konstruktiven Rückmeldungen danken. Ein besonderer Dank gilt ebenfalls Dr. Birgit Heppt, Kristin Schotte, Tatjana Taraszow und meinen anderen großartigen Kolleg*innen am Institut zur Qualitätsentwicklung im Bildungswesen, mit denen es in beruflicher wie persönlicher Hinsicht eine Freude ist, in einem Team zu sein.

Ebenfalls möchte ich den Lehrkräften und Schüler*innen sowie den Erheber*innen und Kodierer*innen danken, ohne deren Engagement die KuL-Studie und damit die vorliegende Dissertation nicht möglich gewesen wären.

Nicht zuletzt danke ich meiner Familie und meinen Freund*innen für den Glauben an mich, die motivierenden Worte und die notwendige Ablenkung. Tim Sandweg bin ich von ganzem Herzen dankbar für den gemeinsamen Aufbruch nach Berlin und freue mich auf alle Orte, die noch kommen werden.

Inhaltsverzeichnis

I Einleitung.....	6
II Theoretischer Rahmen der Arbeit und empirischer Forschungsstand	10
2.1 Begriffsbestimmungen	10
2.1.1 Erwartungen und Einschätzungen von Lehrkräften – Zukunftsgerichtete Prognosen vs. Beurteilungen eines Ist-Zustands	10
2.1.2 Akkuratheit, Inakkuratheit und Verzerrungen	11
2.1.3 Erwartungen und Einschätzungen von Lehrkräften und ihre diagnostische Kompetenz	14
2.2 Urteilsbildung – Bildung akkurater, inakkurater und verzerrter Erwartungen und Einschätzungen.....	15
2.2.1 Modelle der Urteilsakkuratheit	16
2.2.2 Modelle der Urteilsbildung.....	18
2.2.3 Empirische Befunde zur (In)Akkuratheit von Leistungserwartungen und -einschätzungen	25
2.2.4 Empirische Befunde zu Verzerrungen von Leistungserwartungen und -einschätzungen....	29
2.2.5 Empirische Befunde zu Einschätzungen der Motivation und des Arbeitsverhaltens.....	33
2.3 Erwartungen von Lehrkräften und die Leistungsentwicklung von Schüler*innen.....	37
2.3.1 Theorie der selbsterfüllenden Prophezeiung inakkurater Lehrkrafterwartungen.....	37
2.3.2 Empirische Befunde zu Erwartungseffekten und ihren Moderatoren	42
2.3.3 Empirische Befunde zum Prozess von Erwartungseffekten	50
III Ziele und Fragestellungen der Arbeit	56
IV Teilstudien der Arbeit.....	62
4.1 Teilstudie 1	62
4.2 Teilstudie 2	62
4.3 Teilstudie 3	63
V Gesamtdiskussion.....	64
5.1 Zusammenfassung der Befunde und Einordnung in den Forschungsstand	64

5.1.1 Urteilsbildung – (In)Akkuratheit von Erwartungen und Einschätzungen sowie Einflüsse von Stereotypen	65
5.1.2 Effekte von Leistungserwartungen und deren Prozesse.....	76
5.2 Grenzen der Arbeit.....	81
5.2.1 Generalisierbarkeit der Befunde	81
5.2.2 Methodische Einschränkungen	83
5.2.3 Ökologische Validität und Erkenntnisse über Kausalität	84
5.3 Perspektiven für die weitere Forschung	85
5.3.1 Unterschiede zwischen Lehrkräften bzw. Klassen	86
5.3.2 Weiterführende Analysen zum Prozess von Erwartungseffekten	87
5.4 Bedeutung der Erkenntnisse für die pädagogische Praxis	90
5.4.1 Erwartungen von Lehrkräften und Bildungsungleichheiten	90
5.4.2 Aus- und Weiterbildung – Streben nach hohen Erwartungen für alle Schüler*innen.....	91
Literatur.....	95
Schlussteil der Arbeit.....	118
Abstract	118
Zusammenfassung.....	119
Teilstudie 1: Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen	121
Teilstudie 2: Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen	145
Teilstudie 3: Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement.....	171

I Einleitung

Das Bildungssystem erfolgreich zu durchlaufen, ist ein wichtiger Grundstein für gesellschaftliche Teilhabe. Wie bedeutsam Lehrkräfte und ihr Handeln für den Bildungserfolg von Kindern und Jugendlichen sind, hat Hattie (2009) mit seiner Synthese von über 800 Metaanalysen eindrucksvoll dargelegt. Dabei sind die Ansprüche daran, was Lehrkräfte alles leisten sollen, vielfältig. Ihre Kernaufgabe besteht darin, anregende Lerngelegenheiten im Unterricht bereitzustellen und die Kinder und Jugendlichen in ihrer Entwicklung zu fördern (Kunina-Habenicht, Decker & Kunter, 2015). Dazu gehört auch, die jungen Menschen individuell in ihrem Lernen zu unterstützen und ihnen zu ermöglichen, ihre persönlichen Potenziale zu realisieren. Unterricht wird dabei als Angebotsstruktur verstanden, die den Lernenden möglichst gute Anreize für effektive Lernaktivitäten bietet (Seidel & Reiss, 2014). Die Wirkung der Angebote hängt auch davon ab, ob die unterrichtlichen Maßnahmen von den Lernenden wahrgenommen werden und welche kognitiven, motivationalen, emotionalen und volitionalen Prozesse sie bei den Lernenden anstoßen (Angebots-Nutzungs-Modell; Helmke, 2017; Seidel, 2014).

Für die Gestaltung des Unterrichts und anregender Lerngelegenheiten ist auf Seiten der Lehrkraft ihre professionelle Handlungskompetenz zentral. Diese beinhaltet neben Professionswissen, motivationalen Orientierungen und selbstregulativen Fähigkeiten auch Überzeugungen und Werthaltungen, wie etwa Weltbilder, Unterrichtsziele und Erwartungen (Baumert & Kunter, 2006). Überzeugungen von Lehrkräften können sich dabei auf unterschiedliche Ebenen beziehen, etwa auf die Lehrkraft selbst, den Lehr-Lern-Kontext, das Bildungssystem oder die Gesellschaft (Woolfolk Hoy, Davis & Pape, 2006). Überzeugungen zum Lehr-Lern-Kontext beinhalten dabei nicht nur lerntheoretische Überzeugungen, sondern auch Erwartungen an Klassen oder an individuelle Schüler*innen¹. Es gilt als weitgehend belegt, dass Überzeugungen von Lehrkräften eine bedeutende Rolle für ihr Handeln spielen können (vgl. Kunter & Pohlmann, 2015). Welchen Eindruck eine Lehrkraft beispielsweise von einer Schülerin hat, kann ihr Verhalten gegenüber dieser Schülerin auf verschiedene Weise beeinflussen. So kann die Überzeugung der Lehrkraft beispielsweise mitbedingen, wie die Lehrkraft eine Situation mit der Schülerin wahrnimmt, bewertet und wie sie agiert bzw. reagiert (Pajares, 1992). Es liegt daher nahe, dass Erwartungen von Lehrkräften an individuelle Schüler*innen die Gestaltung individueller Lerngelegenheiten mitbestimmen (Schrader & Helmke, 2001). Welche Übungsblätter eine Lehrkraft für einen Schüler auswählt, welche Hilfsmittel

¹ Der vorliegende Text bemüht sich um eine geschlechtergerechte Sprache. Durch die Verwendung von geschlechterneutralen Formen (z. B. Lehrkräfte oder Lernende) sowie des Gendersterns (z. B. Schüler*innen) sollen Personen aller Geschlechter einbezogen werden – explizit auch diejenigen Personen, die sich nicht einem männlichen oder weiblichen Geschlecht zugehörig fühlen.

sie einer Schülerin zur Verfügung stellt, ob sie einem Schüler einen Sachverhalt noch einmal individuell erklärt etc., dürfte bedeutsam davon abhängen, wie die Lehrkraft die Fähigkeiten der Lernenden einschätzt und welche Erwartung sie an das individuelle Lernverhalten hat.

Besonders in Zeiten, in denen vermehrt eine individuelle Förderung jedes einzelnen Kindes gefordert wird, können Erwartungen von Lehrkräften bedeutsam sein. Die Forderung nach Individualisierung ist eng mit der Debatte um heterogene Ausgangsbedingungen von Lernenden und dem Wissen um ethnische, soziale und geschlechtsbezogene Bildungsungleichheiten verknüpft. Spätestens seit der „empirischen Wende“ im Bildungssystem ist bekannt, dass der Bildungserfolg von Kindern und Jugendlichen eng mit ihrer familiären Herkunft und ihrem Geschlecht assoziiert ist (vgl. z. B. Prenzel, Sälzer, Klieme & Köller, 2013; Rjosk, Haag, Heppt & Stanat, 2017). Zur Erklärung dieser Disparitäten werden in der empirischen Bildungsforschung verschiedene Faktoren diskutiert, wie etwa Struktur- und Prozessmerkmale der Familien (z. B. Baumert, Watermann & Schümer, 2003), Sprachkenntnisse der Schüler*innen (z. B. Heinze, Herwartz-Emden & Reiss, 2007), spezifische Sozialisationsprozesse (z. B. Halpern, 2013; Hannover, 2004; Stanat, Bergann & Taraszow, 2018) sowie verschiedene Arten von Diskriminierung (z. B. Diehl & Fick, 2016). Eine weitere Annahme bezieht sich auf Leistungserwartungen von Lehrkräften, die von Stereotypen beeinflusst sein können. Da die Erwartungen eine wichtige Grundlage für das pädagogische Handeln der Lehrkräfte im Unterricht bilden, könnten sie für die Entstehung oder Verfestigung von Bildungsungleichheiten bedeutsam sein.

Im Fokus der drei Teilstudien der vorliegenden Dissertation stehen Erwartungen von Lehrkräften an die Leistungen von Erstklässler*innen. Die ersten beiden Teilstudien befassen sich mit der Frage, auf welche Merkmale der Schüler*innen Lehrkräfte ihre Leistungserwartungen stützen. Ausgehend vom Wissen über ethnische, soziale und geschlechtsbezogene Bildungsungleichheiten untersucht die *erste Teilstudie*, ob und in welchem Ausmaß Leistungserwartungen von Lehrkräften von soziodemografischen Merkmalen der Lernenden beeinflusst sind. Die *zweite Teilstudie* geht dem Ursprung etwaiger Verzerrungen nach und analysiert, ob weitere von der Lehrkraft wahrgenommene Lernressourcen, spezifisch die Motivation und das Arbeitsverhalten der Lernenden, systematische Unterschiede in den Leistungserwartungen von Lehrkräften an verschiedene Gruppen von Schüler*innen erklären können. Die *dritte Teilstudie* befasst sich mit der Bedeutung von Leistungserwartungen für das Unterrichtshandeln der Lehrkräfte sowie für den Lernerfolg der Erstklässler*innen. Dabei wird der Frage nachgegangen, über welche Prozesse Leistungserwartungen von Lehrkräften die Leistungsentwicklung von Lernenden in solch einer Weise beeinflussen, dass die Leistungen letztlich den ursprünglichen Erwartungen entsprechen (*selbsterfüllende Prophezeiung*; Merton, 1948).

Im folgenden *Teil II* der Dissertationsschrift werden die Ziele der Dissertation zunächst theoretisch fundiert und der empirische Forschungsstand beschrieben. Nach einer vorangestellten

Begriffsbestimmung in *Kapitel 2.1*, befasst sich *Kapitel 2.2* mit der Genauigkeit von Leistungserwartungen und -einschätzungen sowie mit einbezogenen Merkmalen der Schüler*innen. Dazu wird zunächst ausgehend von psychologischen Modellen der Urteilsakkuratheit erarbeitet, unter welchen Bedingungen Lehrkräfte zu mehr oder weniger genauen Urteilen gelangen. Daraufhin wird anhand von dualen Prozessmodellen der Urteilsbildung aus der Sozialpsychologie der mögliche Einfluss von Stereotypen beschrieben. Hierbei wird auch herausgearbeitet, dass sich Stereotype auf unterschiedliche Merkmale beziehen und Leistungserwartungen und -einschätzungen von Lehrkräften sowohl direkt als auch indirekt beeinflussen können. Auf die theoretische Einbettung folgt eine Zusammenfassung der empirischen Befundlage zur Genauigkeit von Leistungserwartungen und -einschätzungen von Lehrkräften einerseits und zum Einfluss soziodemografischer Merkmale von Schüler*innen andererseits. Abschließend werden empirische Befunde zur Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens der Lernenden zusammengetragen und deren mögliche Bedeutung für systematische Unterschiede in den Leistungserwartungen für verschiedene Gruppen von Schüler*innen hergeleitet. *Kapitel 2.3* befasst sich mit der Bedeutung von Leistungserwartungen von Lehrkräften für die Leistungsentwicklung von Schüler*innen. Ausgehend vom traditionellen Rosenthal-Experiment, das erstmals für den Schulkontext belegte, dass Erwartungen von Lehrkräften die Leistungsentwicklung von Lernenden in Richtung der Erwartungen beeinflussen können, werden theoretische Annahmen zu Erwartungseffekten im Schulkontext beschrieben. Eine wichtige Annahme bezieht sich darauf, dass sich Erwartungen im Unterrichtshandeln der Lehrkräfte niederschlagen und dadurch den Lernerfolg der Schüler*innen beeinflussen. Nach dieser theoretischen Verortung wird der empirische Forschungsstand in zwei Etappen beschrieben. Zunächst werden aktuelle empirische Befunde zum Auftreten von Erwartungseffekten im Schulkontext sowie zu verstärkenden und abschwächenden Bedingungen zusammengefasst. Anschließend stehen empirische Befunde zu den Erwartungseffekten zugrundeliegenden Prozessen, insbesondere zur Vermittlung über das Lehrkraftverhalten, im Fokus. Auf Basis dieses theoretischen Rahmens und empirischen Forschungsstands werden in *Teil III* die Fragestellungen der vorliegenden Dissertation abgeleitet und die Ziele der Arbeit benannt. In *Teil IV* folgen Zusammenfassungen der drei Teilstudien. *Teil V* widmet sich der Gesamtdiskussion der Erkenntnisse, Grenzen und Implikationen der Dissertation. *Kapitel 5.1* fasst die empirischen Ergebnisse und theoretischen Erkenntnisse der drei Teilstudien zusammen. Zunächst stehen die Ergebnisse und Erkenntnisse zur Genauigkeit von Leistungserwartungen von Lehrkräften, zu ihrer Verzerrung durch soziodemografische Merkmale der Schüler*innen und zur Rolle von Einschätzungen der Motivation und des Arbeitsverhaltens im Fokus (primär Teilstudien 1 und 2). Daraufhin werden die Befunde zur Bedeutung der Erwartungen für die Leistungsentwicklung und zu den zugrundeliegenden Prozessen diskutiert (Teilstudie 3). In *Kapitel 5.2* folgen eine methodische Bewertung und kritische Reflexion der Grenzen der vorliegenden Arbeit, um daraufhin

Perspektiven für zukünftige Forschung (*Kapitel 5.3*) und Implikationen für die pädagogische Praxis (*Kapitel 5.4*) abzuleiten.

II Theoretischer Rahmen der Arbeit und empirischer Forschungsstand

2.1 Begriffsbestimmungen

Im Zentrum der Dissertation stehen Erwartungen von Lehrkräften an zukünftige Leistungen von Schüler*innen. Eng damit verknüpft sind Einschätzungen aktueller Leistungen von Schüler*innen sowie Einschätzungen ihrer Motivation und ihres Arbeitsverhaltens. In Kapitel 2.1.1 werden zunächst Erwartungen und Einschätzungen von Lehrkräften definiert und voneinander abgegrenzt. Ausgehend von der pädagogischen Relevanz von Erwartungen und Einschätzungen, die die tatsächlichen Merkmale der Lernenden möglichst genau widerspiegeln, werden in Kapitel 2.1.2 die Begriffe Akkuratheit, Inakkuratheit und Verzerrungen spezifiziert. In Kapitel 2.1.3 wird der Bezug von Erwartungen und Einschätzungen von Lehrkräften zum nah verwandten Konstrukt der diagnostischen Kompetenz von Lehrkräften erläutert.

2.1.1 Erwartungen und Einschätzungen von Lehrkräften – Zukunftsgerichtete Prognosen vs. Beurteilungen eines Ist-Zustands

Eine explizite theoretische Abgrenzung zwischen Erwartungen und Einschätzungen von Lehrkräften wird nur in einzelnen Forschungsarbeiten vorgenommen (Meissel, Meyer, Yao & Rubie-Davies, 2017; Ready & Chu, 2015; Zhu, Urhahne & Rubie-Davies, 2018). In diesen Arbeiten werden Erwartungen als Prognosen zukünftiger Ausprägungen verstanden, während sich Einschätzungen auf den aktuellen Ist-Zustand beziehen. Leistungsbezogene *Erwartungen* von Lehrkräften sind demzufolge Vorhersagen einer Lehrkraft (z. B. am Anfang eines Schuljahres geäußert) über die *zukünftigen* Leistungen von Schüler*innen (z. B. am Ende eines Schuljahres). Leistungsbezogene *Einschätzungen* von Lehrkräften meinen hingegen Aussagen über die *aktuellen* Leistungen der Schüler*innen zum Zeitpunkt des Urteils.

Wenngleich diese Einteilung logisch erscheint, werden die Begrifflichkeiten in der Forschung eher uneinheitlich genutzt. Die meisten Forscher*innen definieren Erwartungen zwar als zukunftsgerichtete Vorhersagen oder Prognosen von Leistungen oder lernrelevanten Merkmalen von Schüler*innen (z. B. Good, 1987; Good & Nichols, 2001; Jussim, 1986; Peterson, Rubie-Davies, Osborne & Sibley, 2016; Ready & Chu, 2015; Trouilloud, Sarrazin, Martinek & Guillet, 2002; Zhu et al., 2018). Andere Arbeiten fassen konzeptionell aber auch zukunftsgerichtete Prognosen gemeinsam mit Aussagen über den Ist-Zustand unter dem Begriff Erwartungen zusammen (z. B. Brophy & Good, 1974; Hollenstein, Affolter & Brühwiler, 2019; Rubie, 2003). Noch deutlicher wird diese

Uneinheitlichkeit, wenn man die gewählten Operationalisierungen hinzuzieht. Häufig werden zur Messung von als Erwartungen von Lehrkräften bezeichnete Variablen Einschätzungen des Ist-Zustands herangezogen. Lee Jussim beispielsweise erfasste in verschiedenen Studien die Erwartungen von Lehrkräften über auf den Ist-Zustand bezogene Einschätzungen des Talents, der Anstrengung und der Leistung von Lernenden (z. B. Jussim & Eccles, 1992). Gleichzeitig untersuchen solche und ähnliche Studien anhand der Ist-Zustandseinschätzungen der Leistungen von Schüler*innen, ob die Überzeugungen der Lehrkräfte die Leistungsentwicklung der Schüler*innen im Sinne einer selbsterfüllenden Prophezeiung beeinflussen – was auch „Erwartungseffekt“ genannt wird (vgl. Kapitel 2.3). Diese Arbeiten, die Einflüsse auf die Leistungsentwicklung basierend auf Ist-Zustands-Einschätzungen untersuchen, bezeichnen die Leistungsurteile teilweise als Erwartungen von Lehrkräften (bzw. *teacher expectations*; Friedrich, Flunger, Nagengast, Jonkmann & Trautwein, 2015; Hollenstein et al., 2019; Jungbluth, 1994; Jussim & Eccles, 1992; Rubie-Davies & Peterson, 2016; Rubie-Davies et al., 2014; Sorhagen, 2013), teilweise als Einschätzungen von Lehrkräften (bzw. *teacher judgments* oder *teacher perceptions*; Jungbluth, 1993; Ready & Chu, 2015; Robinson-Cimpian, Theule Lubinski, Ganley & Copur-Gencturk, 2014; Stahl, 2007; Zhou & Urhahne, 2013).

In der vorliegenden Arbeit werden Erwartungen und Einschätzungen von Lehrkräften voneinander unterschieden. Mit Erwartungen sind im Folgenden Prognosen über zukünftige Merkmalsausprägungen gemeint. Einschätzungen werden hingegen als Beurteilung des Ist-Zustands verstanden. Beide Begriffe werden in der vorliegenden Arbeit trennscharf zueinander verwendet. In einzelnen Kapiteln wird zur besseren Lesbarkeit teilweise aber auch der Begriff Lehrkrafturteil bzw. Urteil genutzt. Mit diesem Begriff sind Erwartungen und Einschätzungen gleichermaßen gemeint.

2.1.2 Akkuratheit, Inakkuratheit und Verzerrungen

Soll die Bedeutung von Lehrkrafturteilen für Leistungen und Leistungsentwicklungen von Schüler*innen untersucht werden, ist es wichtig die Urteile danach zu unterscheiden, wie sehr sie die tatsächlichen Merkmale der Lernenden widerspiegeln und inwiefern sie von anderen, insbesondere von soziodemografischen Merkmalen der Lernenden beeinflusst sind. Basierend auf den Arbeiten von Jussim, Eccles und Madon (1996) schlagen Ready und Wright (2011) eine Systematisierung vor, die auch in der vorliegenden Arbeit angewendet werden soll. Unterschieden werden Lehrkrafturteile dabei hinsichtlich des Ausmaßes ihrer (In)Akkuratheit (Englisch: *(in)accuracy*) und Verzerrung²

² Für die englischen Begriffe *bias* bzw. *biased* stehen verschiedene Übersetzungen zur Verfügung. In der vorliegenden Dissertation wurde *Verzerrung* bzw. *verzerrt* gewählt, da diese Begriffe eine unerwünschte

(Englisch: *bias*). Die *Akkuratheit* bzw. *Inakkuratheit*³ von Lehrkrafturteilen bezieht sich darauf, wie sehr die Urteile tatsächliche Merkmale der Lernenden widerspiegeln. Konzeptionell bedeutet dies, einen Abgleich zwischen Urteil und Realität vorzunehmen (vgl. Hoffmann, 2017; Schrader & Helmke, 1987). Methodisch nähert man sich diesem Ziel an, indem das subjektive Lehrkrafturteil mit einem möglichst objektiven Maß des Merkmales der Lernenden verglichen wird. So werden Einschätzungen der Leistungen von Schüler*innen beispielsweise dann als akkurat eingestuft, wenn sie mit den Leistungen der Lernenden, gemessen mit einem standardisierten Leistungstest, korrespondieren (vgl. Hoffmann, 2017; Ready & Wright, 2011; Schrader & Helmke, 1987). Abweichungen zwischen den Lehrkrafturteilen und den gemessenen Merkmalen der Lernenden kennzeichnen demnach Inakkuratheit. Das Ausmaß mit dem die Lehrkraft also beispielsweise die Leistungen der individuellen Lernenden über- oder unterschätzt entspricht dem Ausmaß der Inakkuratheit der Einschätzung. Treten solche Über- und Unterschätzungen systematisch für verschiedene Gruppen von Schüler*innen (z. B. in Abhängigkeit soziodemografischer Merkmale wie ihrem ethnischen oder sozialen Hintergrund oder Geschlecht) auf, ist das Lehrkrafturteil in Abhängigkeit des Gruppierungsmerkmals *verzerrt* (vgl. Ready & Wright, 2011: „inaccurate and biased“). Wenn eine Lehrkraft also beispielsweise Kinder aus sozial bessergestellten Familien überdurchschnittlich häufig überschätzt und damit größere soziale Leistungsunterschiede annimmt als standardisierten Testergebnissen zufolge existieren, wären die Lehrkrafturteile in Abhängigkeit vom sozialen Hintergrund der Lernenden verzerrt. Entsprechend werden Lehrkrafturteile nur dann als „verzerrt“ bezeichnet, wenn sie Unterschiede zwischen Gruppen von Schüler*innen über- oder unterschätzen (vgl. Ready & Wright, 2011; siehe auch Jussim, McCauley & Lee, 1995).

Dem herangezogenen Vergleichsmaßstab oder Kriterium kommt bei der Bestimmung des Ausmaßes der (In)Akkuratheit eine zentrale Bedeutung zu. Bei Ist-Zustandseinschätzungen stellt ein objektives Maß der derzeitigen Merkmalsausprägung den Vergleichsmaßstab dar, also beispielsweise bei Einschätzungen der derzeitigen Leistungen die Ergebnisse in einem zeitnah durchgeführten Leistungstest. Die Güte von Prognosen lässt sich hingegen streng genommen nur im Längsschnitt beurteilen. Hierzu wird die auf der Grundlage vorliegender Informationen gebildete Erwartung mit dem in der Zukunft eingetretenen Ergebnis abgeglichen. Es muss also zum Beispiel überprüft werden, ob die Prognose einer Lehrkraft über die Endjahresleistung einer Schülerin, die die Lehrkraft zu Beginn des Schuljahrs formuliert hat, mit der letztlich in einem Leistungstest erzielten Endjahresleistung der Schülerin übereinstimmt. Diese Art der Akkuratheit von Erwartungen wird als

Beeinflussung zum Ausdruck bringen, aber gleichzeitig die Quelle des Einflusses offenlassen (anders als z. B. *vorurteilsbehaftet*).

³ Abweichend von dieser Begrifflichkeit wurde in Teilstudie 1 von zutreffenden bzw. unzutreffenden Erwartungen gesprochen. Gemeint sind damit akkurate bzw. inakkurate Erwartungen.

Akkuratheit der Vorhersage (*predictive accuracy*) bezeichnet (vgl. Jussim, 1991; Trouilloud et al., 2002). Dabei ist allerdings zu berücksichtigen, dass die Erwartung nur dann als akkurat definiert wird, wenn sie die Endjahresleistung zwar zutreffend vorhergesagt, diese aber nicht – zum Beispiel über die Prozesse einer selbsterfüllenden Prophezeiung – verursacht hat (*prediction without causation*; z. B. Jussim & Harber, 2005).

Von dieser längsschnittlichen Perspektive auf die Akkuratheit von Erwartungen wird ein zweiter Akkuratheitsaspekt – die Akkuratheit des Eindrucks (*impression accuracy*; vgl. Jussim, 1991; Trouilloud et al., 2002) – abgegrenzt, der sich auf den Zeitpunkt der Erwartungsbildung bezieht und damit dem Verständnis von Akkuratheit bei Einschätzungen ähnelt. Die Akkuratheit des Eindrucks beschreibt, ob aus den zum Zeitpunkt der Erwartungsbildung vorliegenden Informationen nur valide und keine invaliden Informationen in die Erwartung einbezogen werden. Erwartungen, die auf validen Informationen basieren, werden dabei als akkurater eingeordnet als Erwartungen, die auf weniger validen Informationen basieren. Welche Informationen jedoch als valide und angemessene Basis von Leistungserwartungen bewertet werden, variiert zwischen den Forschungsarbeiten. Viele Forschende legen ihren Arbeiten ein eher enges Verständnis von Akkuratheit in Leistungserwartungen zugrunde (z. B. Hinnant, O'Brien & Ghazarian, 2009; McKown & Weinstein, 2002; Rubie-Davies & Peterson, 2016). Ihr Vorgehen legt nahe, dass Erwartungen ausschließlich auf den aktuellen, tatsächlich beobachtbaren Leistungen der Schüler*innen basieren sollten. Andere, insbesondere neuere Forschungsarbeiten vertreten hingegen ein breiteres Verständnis von Akkuratheit bei Erwartungen (z. B. Becker, 2013; de Boer, Bosker & van der Werf, 2010; Hollenstein et al., 2019). Diese Arbeiten orientieren sich an der Perspektive der pädagogischen Praxis, aus der heraus es durchaus plausibel ist, weitere lernrelevante Merkmale wie etwa die Motivation oder Anstrengungsbereitschaft von Lernenden in Prognosen zukünftiger Leistungen einzubeziehen, da diese Merkmale als wichtige Bedingungsfaktoren für Leistungsentwicklungen bekannt sind (vgl. z. B. Spinath, 2011). Weitere leistungsfernere Merkmale, wie etwa der soziale Hintergrund von Lernenden, ein Zuwanderungshintergrund oder ihr Geschlecht, sollten in einem meritokratischen Schulsystem hingegen keine Rolle für Leistungsurteile von Lehrkräften spielen – weder für Einschätzungen noch Erwartungen (vgl. de Boer et al., 2010).

Die ersten beiden Teilstudien der vorliegenden Dissertation befassen sich mit der Bildung von Erwartungen und Einschätzungen und dem Einbezug der zum Zeitpunkt der Urteilsbildung vorliegenden Informationen (im Sinne der *impression accuracy*). Wird also im Folgenden von Akkuratheit gesprochen, ist die Akkuratheit auf Basis der zum Zeitpunkt der Urteilsbildung vorliegenden Informationen und nicht der Vergleich mit einem in Zukunft eingetretenen Ergebnis

gemeint.⁴ Dabei folgt die vorliegende Arbeit dem breiteren Verständnis von Akkuratheit. Entsprechend werden Leistungserwartungen dann als akkurat definiert, wenn sie mit vielfältigen Maßen für die aktuellen Leistungen der Schüler*innen, ihren allgemeinen kognitiven Fähigkeiten sowie ihrer Motivation und Anstrengungsbereitschaft kongruieren. Bei Einschätzungen von aktuellen Merkmalen der Lernenden werden nur die jeweiligen Merkmale selbst als Vergleichsmaßstab herangezogen.

2.1.3 Erwartungen und Einschätzungen von Lehrkräften und ihre diagnostische Kompetenz

Ein mit Erwartungen und Einschätzungen von Lehrkräften nah verwandtes Konstrukt ist ihre diagnostische Kompetenz. Während sich Erwartungen und Einschätzungen an individuelle Schüler*innen innerhalb der professionellen Handlungskompetenz von Lehrkräften (Baumert & Kunter, 2006) dem Bereich der Werthaltungen und Überzeugungen zuordnen lassen (Keller, 2016; Kunina-Habenicht et al., 2015), wird die diagnostische Kompetenz von Lehrkräften dem Bereich des Professionswissens, spezifischer dem pädagogischen Wissen, zugeteilt (Baumert & Kunter, 2006; Praetorius & Südkamp, 2017).

Unter *diagnostischer Kompetenz* von Lehrkräften wird im engeren Sinn die Fähigkeit verstanden, Merkmale von Personen zutreffend einzuschätzen (Schrader, 2010). In einem weiteren Verständnis werden auch Urteile über Merkmale didaktischer Materialien (z. B. Aufgaben) einbezogen (McElvany et al., 2009; Südkamp, Möller & Pohlmann, 2008). Diesem weiteren Verständnis folgend differenzieren McElvany et al. (2009) verschiedene Urteilsgegenstände anhand von zwei übergeordneten Dimensionen: der *Urteilsbereich* (Schüler*innen vs. Unterrichtsmaterialien) und die *Urteilebene*, die von sehr spezifisch – wie etwa individuelle Schüler*innen bzw. konkrete Aufgaben – zu eher global – wie beispielsweise eine Klassenstufe bzw. ein Themenbereich – reicht. Auf jeder der Kombinationen aus Urteilsbereich und Urteilebene kann wiederum die von Schrader und Kollegen vorgeschlagene (Schrader, 1989; Schrader & Helmke, 1987) und mittlerweile sehr weit verbreitete Einteilung diagnostischer Kompetenz in drei voneinander weitgehend unabhängige Komponenten (Spinath, 2005) erfolgen: die Niveau-, Differenzierungs- und Rang- oder Vergleichskomponente. Die *Niveauebene* bezieht sich auf Einschätzungen des absoluten Fähigkeits- oder Schwierigkeitslevels. Die *Differenzierungskomponente* umfasst Einschätzungen der

⁴ Teilstudie 3 befasst sich mit selbsterfüllenden Prophezeiungen von Leistungserwartungen. Da sich die Akkuratheit der Vorhersage (*predictive accuracy*) statistisch als Differenz aus der bivariaten Korrelation zwischen Erwartungen und Endjahresleistung und dem standardisierten Effekt der selbsterfüllenden Prophezeiung definieren lässt (z. B. Jussim & Harber, 2005), wird dieser Aspekt von Akkuratheit in der vorliegenden Dissertation ebenfalls berührt, steht aber nicht im Zentrum.

Variation oder Streuung des jeweiligen Merkmals. Die *Vergleichs- oder Rangkomponente* beschreibt Einschätzungen über die Rangfolge der Merkmalsausprägung in einer Gruppe von Personen bzw. bei mehreren Aufgaben. Bei Untersuchungen der diagnostischen Kompetenz von Lehrkräften wird zumeist die Rangkomponente herangezogen und von einigen Autor*innen auch als der zentrale Indikator diagnostischer Kompetenz angesehen (*diagnostische Sensitivität*; Schrader & Helmke, 1987; Weinert & Schrader, 1986).

Wie hängt die diagnostische Kompetenz von Lehrkräften mit ihren Erwartungen und Einschätzungen sowie deren Akkuratheit zusammen? Die obige Definition von diagnostischer Kompetenz als Fähigkeit, Merkmale zutreffend einzuschätzen, verdeutlicht bereits, dass akkurate Urteile als Ausdruck einer hohen diagnostischen Kompetenz der Lehrkräfte verstanden werden können. Man würde also annehmen, dass die diagnostische Kompetenz von Lehrkräften als zugrundeliegende Fähigkeit die individuellen Urteile über Schüler*innen beeinflusst. Sie dürfte sich also u. a. in mehr oder weniger akkuraten Einschätzungen der Leistung oder der Motivation und des Arbeitsverhaltens von Schüler*innen niederschlagen und auf diesem Weg auch die Erwartungen an zukünftige Leistungen beeinflussen (Südkamp, Kaiser & Möller, 2012). Konzeptionell ist die diagnostische Kompetenz also auf der Ebene der Lehrkraft angesiedelt, wohingegen die Akkuratheit von Lehrkrafturteilen auf der Ebene der einzelnen Einschätzungen und Erwartungen, also letztlich auf der Ebene der individuellen Schüler*innen, angesiedelt ist. Diese Hierarchie zeigt sich auch in der Operationalisierung von diagnostischer Kompetenz. Um die diagnostische Kompetenz einer Lehrkraft zu bestimmen, wird zumeist ein aggregiertes Maß der Akkuratheit über die Einschätzungen vieler individueller Schüler*innen hinweg bestimmt. Die diagnostische Kompetenz einer Lehrkraft wird demzufolge zumeist als mittlere Akkuratheit der einzelnen, vorgenommenen Urteile definiert.

2.2 Urteilsbildung – Bildung akkurater, inakkurater und verzerrter Erwartungen und Einschätzungen

Nachdem Erwartungen und Einschätzungen von Lehrkräften nun begrifflich und konzeptionell eingeordnet und mit dem verwandten Konstrukt der diagnostischen Kompetenz von Lehrkräften in Bezug gesetzt wurden, widmet sich Kapitel 2.2 der Frage, weshalb und unter welchen Bedingungen Urteile von Lehrkräften mehr oder weniger akkurat ausfallen und teilweise in Abhängigkeit von soziodemografischen Merkmalen der Schüler*innen verzerrt sein können. Als theoretische Basis für die erste Frage sind psychologische Modelle der Urteilsakkuratheit nützlich. Weit verbreitet ist das Linsenmodell von Brunswik (1956), das beschreibt, wie Personen zu Urteilen über nicht direkt beobachtbare Merkmale gelangen. Dieses Modell soll im folgenden Kapitel 2.2.1 erläutert werden.

Zum Verständnis, wie systematische Verzerrungen von Urteilen entstehen, dienen duale Prozessmodelle der Urteilsbildung aus der Sozialpsychologie. Diese werden in Kapitel 2.2.2 beschrieben.

2.2.1 Modelle der Urteilsakkuratheit

Das *Linsenmodell* (Brunswik, 1956) ist ein allgemeines psychologisches Rahmenmodell, welches genutzt werden kann, um die einer Urteilsbildung zugrundeliegenden Prozesse zu untersuchen und um zu verstehen, ob, wann und warum Urteile mehr oder weniger akkurat ausfallen (vgl. Förster & Böhmer, 2017; Nestler & Back, 2013). Das Modell (vgl. Abb. 1) nimmt an, dass die urteilende Person einen Fokus – wie durch eine Linse – auf eine Auswahl beobachtbarer Merkmale setzt, um auf ein nicht direkt beobachtbares Merkmal der einzuschätzenden Person zu schließen. Ein solches nicht direkt beobachtbares Merkmal könnte im Schulkontext beispielsweise die Rechenfähigkeit einer Schülerin (Annika) sein (vgl. Förster & Böhmer, 2017, die das Leseverständnis eines Schülers als Beispiel heranziehen). Da die Rechenfähigkeit von Annika zunächst nicht direkt beobachtbar ist, wird ihre Lehrerin beobachtbare Merkmale heranziehen, die ihrer Meinung nach Rückschlüsse auf Annikas Rechenfähigkeit zulassen. So könnte sie beispielsweise beobachten, dass Annika sehr flüssig und weitgehend fehlerfrei bis 50 zählt, Bonbons gleichmäßig auf sich und ihre Freundinnen aufteilt und sich häufig mit mathematischen Problemen beschäftigt. Außerdem fällt der Lehrerin auf, wie ehrgeizig Annika im Unterricht insgesamt mitarbeitet und wie ordentlich sie ihre Schulmaterialien behandelt. Möglicherweise weiß die Lehrerin zusätzlich noch über Annika, dass sie beim letzten Sportfest den ersten Platz belegt hat, ihre Mutter Ärztin und ihr Vater Richter ist. Durch diese und ähnliche Beobachtungen kann die Lehrerin zu der Einschätzung gelangen, dass Annika eine recht hohe Rechenfähigkeit aufweist. Wird Annikas Rechenfähigkeit zusätzlich anhand eines standardisierten Rechentests erfasst, lässt sich die Akkuratheit der Einschätzung der Lehrerin bestimmen. Das Linsenmodell unterscheidet dazu drei Variablenbereiche (Back & Nestler, 2016; Förster & Böhmer, 2017; Nestler & Back, 2013): das tatsächliche Merkmal der Person (Annikas tatsächliche Rechenfähigkeit), das Urteil über das Merkmal der Person (die Einschätzung der Lehrerin zu Annikas Rechenfähigkeit) und die in einer Situation beobachtbaren Signale (z. B. Annikas Zählflüssigkeit bis 50 oder die ordentlichen Schulmaterialien). Die Beziehung dieser drei Variablenbereiche bedingt das Ausmaß der Urteilsakkuratheit (vgl. Nestler & Back, 2013). Zunächst hängt die Akkuratheit davon ab, wie valide die beobachtbaren Signale für das nicht direkt beobachtbare Merkmal sind (*Signalvalidität*). Zudem variiert, inwiefern die urteilende Person ein spezielles Signal für ihr Urteil

heranzieht (*Signalnutzung*). Die Signalsensitivität beschreibt schließlich, ob die beurteilende Person valide und keine invaliden Signale für die Einschätzung heranzieht (*Sensitivität*).

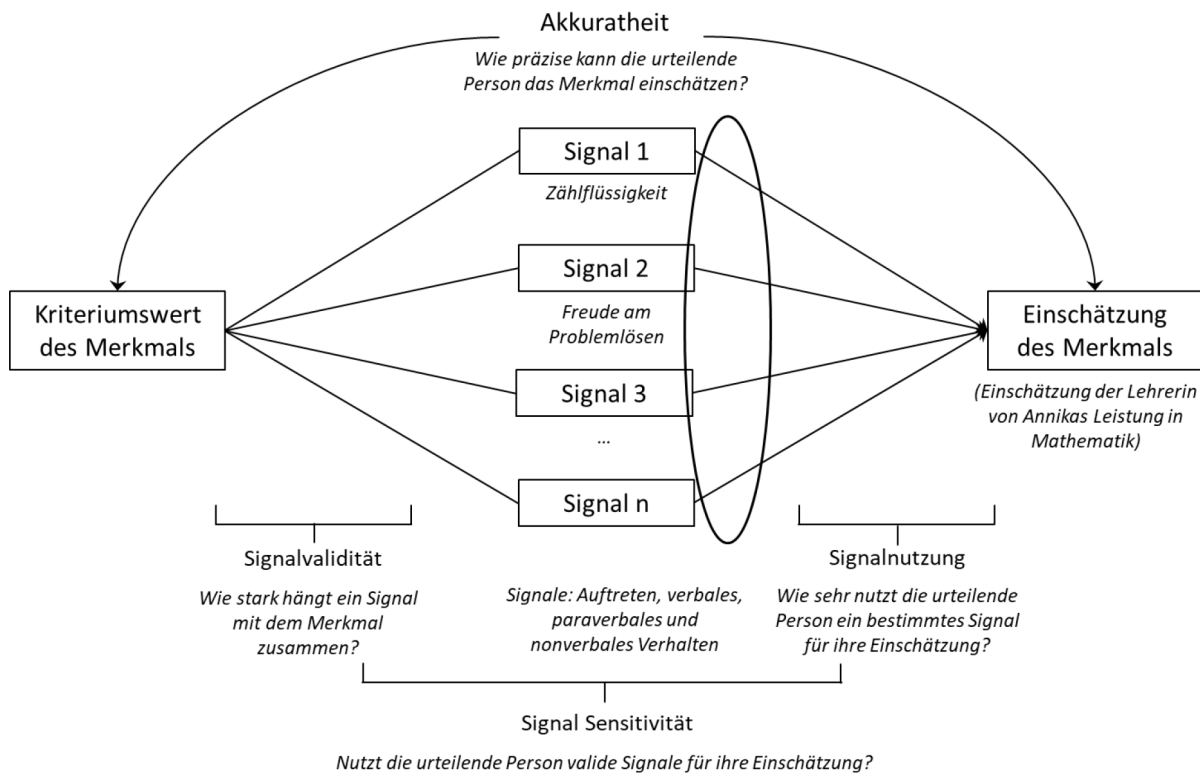


Abb. 1: Linsenmodell (adaptiert nach Back & Nestler, 2016; Förster & Böhmer, 2017).

Auf das obige Beispiel angewendet, bezieht sich die Signalvalidität auf das Ausmaß, mit dem die beobachteten Signale, also z. B. die Zähflüssigkeit bis 50, das gleichmäßige Aufteilen von Bonbons auf Freunde und die Freude am Lösen mathematischer Probleme mit der Rechenfähigkeit einer Person zusammenhängen. Nehmen wir einmal an, dass die Zähflüssigkeit und das gleichmäßige Aufteilen von Bonbons gute Indikatoren für Rechenfähigkeit sind, wohingegen die Freude am Lösen mathematischer Probleme nur bedingt mit der Rechenfähigkeit zusammenhängt. Würde Annikas Lehrerin nun vor allem die Zähflüssigkeit und das gleichmäßige Aufteilen der Bonbons als valide Informationen, nicht aber die Freude am Lösen mathematischer Probleme oder den Umstand, dass Annikas Mutter Ärztin und ihr Vater Richter ist (invalide Information), für ihre Einschätzung (Annika hat hohe Fähigkeiten im Rechnen) heranziehen (Signalnutzung), würde dies eine hohe Sensitivität der Lehrkraft kennzeichnen. Da in dieser beispielhaften Situation valide Signale für Annikas Rechenfähigkeit vorlagen und die Lehrkraft diese validen Signale nicht aber invalide Signale genutzt hat, dürfte die resultierende Einschätzung relativ akkurat ausfallen. Würde die Lehrkraft ihre Einschätzung hingegen vor allem auf die Freude am Lösen mathematischer Probleme oder die Berufe von Annikas Eltern stützten – entweder weil nur diese Informationen verfügbar sind oder weil die

Lehrkraft annimmt, diese Informationen seien valide, – sollte daraus eine weniger akkurate Einschätzung resultieren.

Die Anwendung des Linsenmodells der Urteilsakkuratheit auf das Beispiel eines mathematischen Leistungsurteils zeigt, dass verschiedene Aspekte beeinflussen können, wie genau das Urteil der Lehrkraft ausfällt. Um zu verstehen warum bei der Urteilsbildung möglicherweise auch invalide Signale genutzt werden und wie es zu verzerrten Urteilen kommen kann, sind sozialpsychologische Modelle der Urteilsbildung hilfreich.

2.2.2 Modelle der Urteilsbildung

Sozialpsychologische Modelle der Urteilsbildung beschreiben, wie Personen in sozialen Situationen Informationen verarbeiten und auf dieser Basis zu Urteilen über andere Personen gelangen. Am weitesten verbreitet sind duale Prozessmodelle, die im Kern davon ausgehen, dass sich die Informationsverarbeitung in sozialen Situationen zwischen den beiden Polen der automatischen, Ressourcen schonenden kategorienbasierten Verarbeitung (z. B. *category-based*: Fiske & Neuberg, 1990; *heuristic*: Ferreira, Garcia-Marques, Sherman & Sherman, 2006) und der bewusst erfolgenden, ressourcenintensiven eigenschaftsbasierten Verarbeitung (z. B. *information integrative*: Fiske & Neuberg, 1990; *rule-based*: Ferreira et al., 2006) bewegt (vgl. auch Böhmer, M., Englich & Böhmer, 2017). Eins der bekanntesten Modelle ist das *Kontinuum-Modell der Eindrucksbildung* von Fiske und Kolleg*innen (Fiske, Lin & Neuberg, 1999; Fiske & Neuberg, 1990). Das Modell (vgl. Abb. 2) geht davon aus, dass sich Menschen in sozialen Situationen in unterschiedlich intensiven Verarbeitungsprozessen ein Bild von ihrem Gegenüber machen. Die Intensität der Verarbeitung bewegt sich dabei auf einer Art Kontinuum von kategorienbasierter hin zu eigenschaftsbasierter Informationsverarbeitung. Dem Modell zufolge findet bei jeder sozialen Begegnung zunächst eine Kategorisierung der Person anhand eines leicht zu identifizierenden, salienten Merkmals statt. Zu solchen Merkmalen gehören vorzugsweise visuell verankerte Merkmale wie Körperform oder Hautfarbe, aber auch sozial bedeutsame Kategorien wie Geschlecht, Alter oder sozialer Status (vgl. z. B. Fiske & Neuberg, 1990). Es können aber auch weitere Merkmale zur Kategorisierung dienen, die leicht zugänglich sind. Dieser kategorisierende Prozess ist dadurch gekennzeichnet, dass er automatisch und unter Inanspruchnahme minimaler zeitlicher und kognitiver Ressourcen erfolgt. Damit dient die Kategorisierung einer schnellen und effizienten mentalen Organisation unserer Umwelt, die in vielen Situationen bereits ausreicht.

Ob nach diesem ersten, unbewusst und schnell ablaufenden Prozess der Kategorisierung (z. B. „Mädchen“) nach weiteren Informationen über die zu beurteilende Person gesucht wird, hängt von

verschiedenen Bedingungen ab. Hierzu gehören die *Verfügbarkeit kognitiver und zeitlicher Ressourcen* sowie die *Motivation zur weiteren Informationensuche* auf Seiten der beurteilenden Person genauso wie die *Verfügbarkeit und Art weiterer Informationen* über die zu beurteilende Person. Wenn diese Voraussetzungen erfüllt sind, kann eine weiterführende Informationensuche und -verarbeitung stattfinden. Hierbei werden dem Kontinuum-Modell zufolge neue Informationen mit der ursprünglich zugewiesenen Kategorie abgeglichen. Bewertet die beurteilende Person die neu hinzugezogenen Merkmale als kongruent zur Kategorie (und gibt es sonst keinen Grund aufwändig nach weiteren Informationen zu suchen) ist der Prozess der Eindrucksbildung abgeschlossen (*category confirmation*). Die beurteilende Person zeigt dann affektive, kognitive und verhaltensbezogene Tendenzen gegenüber der beurteilten Person, die mit dieser Kategorisierung verknüpft sind. Was diese aktivierten Tendenzen für die Bildung von Erwartungen und Einschätzungen von Lehrkräften bedeuten können, soll später in diesem Kapitel thematisiert werden.

Bewertet die beurteilende Person die neu hinzugezogenen Merkmale der Person hingegen als inkongruent zur Kategorie, wird eine *Rekategorisierung* wahrscheinlich, in der diese neuen Merkmale integriert werden (z. B. „naturwissenschaftlich interessiertes Mädchen“). Wie stark sich die Informationsverarbeitung im Folgenden hin zu einer vollständig eigenschaftsbasierten Verarbeitung verschiebt, hängt wiederum von der Motivation, den zeitlichen und kognitiven Ressourcen der beurteilenden Person, von der Art und Verfügbarkeit weiterer Informationen sowie von der wahrgenommenen Passung zwischen den neuen Merkmalen und der jeweils aktuellen Kategorisierung ab. Die am stärksten eigenschaftsbasierte Informationsverarbeitung (*piecemeal integration*, Fiske et al., 1999) ist dem Modell zufolge dadurch gekennzeichnet, dass alle relevanten Merkmale Stück-für-Stück in den gebildeten Eindruck integriert werden und zum letztlichen Gesamteindruck beitragen. Die ursprüngliche Kategorisierung verschwindet dabei nicht einfach, sondern wird zu einem Merkmal der zu beurteilenden Person unter vielen. Eine solche stark eigenschaftsbasierte Informationsverarbeitung ist sehr ressourcenintensiv und kommt Fiske et al. (1999) zufolge nur relativ selten vor. Häufiger findet eine Informationsverarbeitung statt, die sich zwischen den beiden Polen der vollständig kategorienbasierten und der vollständig eigenschaftsbasierten Verarbeitung bewegt.

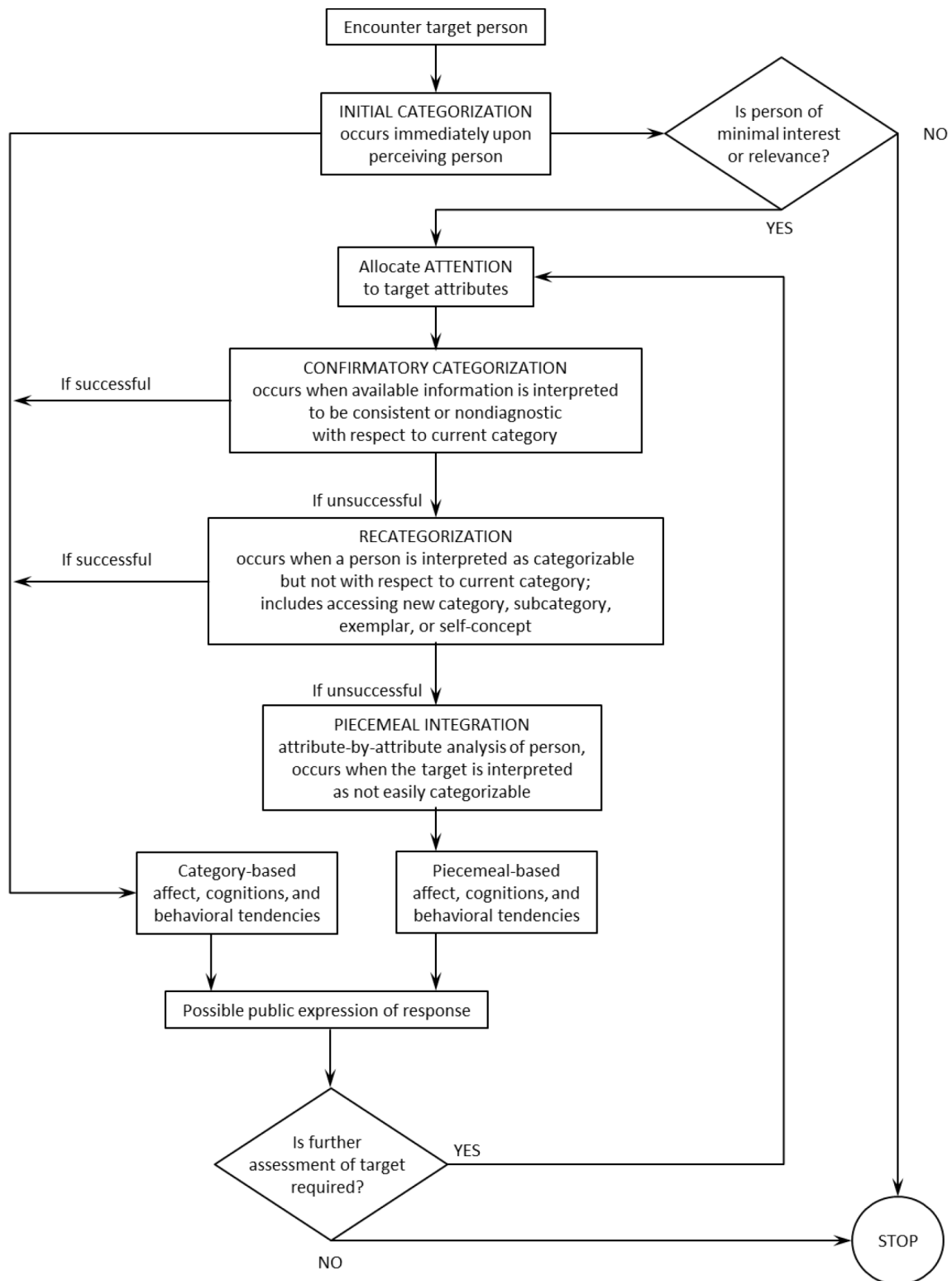


Abb. 2: Kontinuum-Modell der Eindrucksbildung (Fiske, 1986; zitiert nach Fiske et al., 1999).

Es wurde bereits angedeutet, dass von verschiedenen Bedingungen abhängt, wie stark sich die Verarbeitung auf dem Kontinuum hin zur vollständig eigenschaftsbasierten Verarbeitung verschiebt bzw. wie stark die Eindrucksbildung von Kategorisierungen gekennzeichnet bleibt. Dabei ist eine Verschiebung in Richtung eigenschaftsbasierter Informationsverarbeitung generell nur dann möglich, wenn ausreichend zeitliche und kognitive Ressourcen auf Seiten der beurteilenden Person zur Verfügung stehen. Darüber hinaus spielen die *Motivation zur weiteren Informationssuche* sowie die *Verfügbarkeit und Art weiterer Informationen* bei Verschiebungen auf dem Kontinuum eine Rolle. Für diese Bedingungsfaktoren soll jeweils kurz erläutert werden, wie sie zu einer eigenschaftsbasierten Eindrucksbildung beitragen, und welche Rolle ihnen im Schulkontext zukommen könnte.

Im Bereich der Motivation sind zunächst die Hauptmotive – Streben nach Zugehörigkeit, Verstehen, Kontrolle, Selbstwert und Vertrauen – zu nennen. Wenn diese Hauptmotive in einer Situation stark ausgeprägt sind, wird eine eigenschaftsbasierte Informationsverarbeitung wahrscheinlicher (Fiske et al., 1999). Wie genau sich diese Hauptmotive in der Informationsverarbeitung niederschlagen, hängt dabei substantiell mit individuellen Bedürfnissen, Werthaltungen und Zielen zusammen. Ist eine Person beispielsweise davon überzeugt, dass man allen Menschen möglichst unvoreingenommen gegenüberzutreten sollte, wäre eine stärker eigenschaftsbasierte Informationsverarbeitung selbstwertdienlich, weil sie der Überzeugung entspricht. Die Person würde daher versuchen, möglichst viele individuelle Information über ihr Gegenüber zu sammeln und ihren eigenen Selbstwert dadurch aufrechtzuerhalten oder zu steigern (vgl. Fiske et al., 1999). Dass solche Prozesse auch in der Schule relevant sind, ist gut vorstellbar. So kann angenommen werden, dass das Streben nach objektiven Urteilen zum beruflichen Ethos von Lehrkräften gehört und die meisten Lehrkräfte entsprechend die Wertvorstellung teilen, die Kinder soweit möglich in ihren individuellen Fähigkeiten korrekt einzuschätzen. Basierend auf dieser Wertvorstellung würde sich eine Lehrkraft selbstwertdienlich verhalten, wenn sie sich um möglichst akkurate Urteile bemüht.

Zudem erhöht eine antizipierte Rechenschaftspflicht die Motivation, zu möglichst zutreffenden Einschätzungen zu gelangen (z. B. Hattrup & Ford, 1995; Pendry & Macrae, 1996; Tetlock, 1992; Tetlock & Lerner, 1999). Eine solche antizipierte Rechenschaftspflicht kann an verschiedenen Stellen der Schullaufbahn auftreten, beispielsweise wenn Lehrkräfte Noten oder Empfehlungen für die weitere Schullaufbahn eines Kindes geben. Bei solchen Urteilen ist es gängig, dass Elterngespräche stattfinden oder Rücksprachen mit Kolleg*innen oder mit der Schulleitung vorgesehen sind. In der alltäglichen Interaktion dürften Lehrkräfte allerdings weitaus seltener annehmen, sie müssten für eine Erwartung oder Einschätzung Rechenschaft ablegen.

Gleichzeitig ist aus Experimenten (z. B. Berscheid, Graziano, Monson & Dermer, 1976) bekannt, dass die reine Erwartung einer längerfristigen Zusammenarbeit mit einer Person (*committed outcome dependency*; Fiske & Neuberg, 1990) eine stärker individualisierende Informationsverarbeitung begünstigen kann. Dies wird damit erklärt, dass wir mit Personen, mit denen wir längerfristiger in Kontakt stehen werden, ein positives Interaktionsklima anstreben. Um dies zu erreichen ist es nützlich, sein Gegenüber möglichst gut zu kennen und zutreffend einzuschätzen. Im Schulkontext wird die Erwartung einer längerfristigen Zusammenarbeit häufig zutreffen, da Lehrkräfte typischerweise für das gesamte Schuljahr, häufig auch für mehrere Jahre, eine Klasse unterrichten.

Um stärker eigenschaftsbasierte Verarbeitungsprozesse überhaupt zu ermöglichen, müssen zunächst *individuelle Informationen verfügbar* sein (Fiske et al., 1999). Insbesondere dann, wenn einer Person nur wenige relevante Informationen über eine andere Person vorliegen, wird der Einfluss von einfach wahrnehmbaren Gruppierungsmerkmalen und damit der kategorienbasierten Verarbeitung stärker. Im Schulkontext liegen vor allem dann noch wenige Informationen über einzelne Schüler*innen vor, wenn die Lehrkraft die Kinder erst seit kurzem kennt, beispielsweise weil eine Klasse gerade neu zusammengesetzt wurde. In dieser Situation dürfte der kategorienbasierten Eindrucksbildung daher stärkeres Gewicht zukommen als später im Schuljahr, wenn die Lehrkraft auf vielfältige Erfahrungen mit den Schüler*innen zurückgreifen kann.

Verschiebungen auf dem Kontinuum zwischen kategorienbasierter und eigenschaftsbasierter Informationsverarbeitung hängen weiterhin vom *wahrgenommenen Maß der Kategorienkonsistenz* der verfügbaren Informationen ab (vgl. z. B. Fiske et al., 1999; Fiske & Neuberg, 1990). Wie bereits beschrieben wird eine weitere Informationssuche nur dann angestoßen, wenn neue Informationen von der beurteilenden Person als nicht übereinstimmend mit der aktuellen Kategorie bewertet werden. In diesem Bewertungsprozess spielt die Zugehörigkeit der beurteilenden Person zu einer Gruppe eine wichtige Rolle. Personen einer Fremdgruppe werden generell homogener wahrgenommen als Individuen der eigenen Gruppe, was dazu führt, dass individuelle Informationen über Personen der Eigengruppe leichter verarbeitet werden können (Fiske & Taylor, 1991). Daher ist eine stärker eigenschaftsbasierte Informationsverarbeitung für Individuen der eigenen Gruppe generell wahrscheinlicher. Für Mitglieder einer Fremdgruppe wird hingegen eine vorgenommene Kategorisierung eher als passend angenommen und daher beibehalten. Da Lehrkräfte in Grundschulen in Deutschland primär weiblich sind und keinen Zuwanderungshintergrund haben, wäre es denkbar, dass neue Informationen für Jungen und für Kinder mit Zuwanderungshintergrund weniger differenziert wahrgenommen und daher wahrscheinlicher als kategorien-konsistent beurteilt werden. Die Informationssuche dürfte dann für diese Kinder etwas schneller beendet werden. Kategorienbasierte Einschätzungen könnten daher für Jungen und für Kinder mit

Zuwanderungshintergrund etwas wahrscheinlicher sein als für Mädchen und für Kinder ohne Zuwanderungshintergrund.

Bedeutung für Akkuratheit, Inakkuratheit und Verzerrungen von Erwartungen und Einschätzungen

Was bedeuten die beiden Modi der Informationsverarbeitung – kategorienbasiert und eigenschaftsbasiert – für das Ausmaß von Akkuratheit, Inakkuratheit und Verzerrungen in Erwartungen und Einschätzungen von Lehrkräften? Unbedacht läge die Vermutung nahe, dass eine kategorienbasierte Verarbeitung zu inakkuraten und nach dem Gruppierungsmerkmal verzerrten Erwartungen und Einschätzungen führt und eine eigenschaftsbasierte Verarbeitung stets in akkuraten Urteilen mündet. Diese Vermutung wird der Komplexität allerdings nur bedingt gerecht und ist zudem nicht in allen Fällen zutreffend. Um dies nachvollziehen zu können, ist es zunächst wichtig zu verstehen, welche Bedeutung eine kategorienbasierte Verarbeitung für die Erwartungen und Einschätzungen einer Person hat.

Bei der kategorienbasierten Verarbeitung wird das Gegenüber einer Gruppe von Menschen zugeordnet. Die Emotionen, Kognitionen und das Verhalten der beurteilenden Person werden dann im Weiteren von der Kategorie bestimmt. An dieser Stelle werden Stereotype relevant, also verallgemeinernde Überzeugungen, Wissen und Erwartungen über die Merkmale von Personen, die einer Gruppe zugehören (Pendry, 2014). Die Zuordnung einer Person zu einer sozialen Gruppe aktiviert die mit dieser Gruppe verbundenen Stereotype. Die Einordnung eines Kindes beispielsweise in die Kategorie Mädchen kann verschiedene Stereotype über die Gruppe der Mädchen aktivieren. So könnte angenommen werden, Mädchen seien zuvorkommend, hilfsbereit, zurückhaltend, schulisch motiviert, gut im Lesen, schwach in Mathematik und so weiter. Solche und ähnliche Annahmen über die Merkmale von Mädchen sind im Gedächtnis in Form kognitiver Netzwerke verankert, die durch die Kategorisierung aktiviert werden. Durch diese Aktivierung sind die stereotypen Annahmen dann kognitiv leichter zugänglich als die Wahrnehmung alternativer Merkmalsausprägungen. Nehmen wir einmal an, eine Mathematiklehrkraft unterrichtet eine Klasse vertretungsweise. Da das Geschlecht eine saliente Kategorie ist, würde die Lehrkraft die Kinder kognitiv schnell in Mädchen und Jungen einteilen. Damit einhergehend kann das Stereotyp, Mädchen seien schwächer in Mathematik als Jungen, aktiviert werden und beeinflussen, wie die Lehrkraft die individuellen Mathematikfähigkeiten der Mädchen und Jungen der Klasse einschätzt.

Welche Auswirkung die Anwendung dieses Stereotyps auf die (In)Akkuratheit und Verzerrung der Leistungserwartungen und -einschätzungen hat, muss anhand zweier Aspekte betrachtet werden. Einerseits stellt sich die Frage, ob das Stereotyp selbst akkurat ist. Andererseits müssen verschiedene

Ebenen der Bewertung unterschieden werden – die Ebene der Gruppe und die des Individuums. Zunächst können Stereotype akkurat oder inakkurat sein (Bodenhausen, Kang & Peery, 2012; Jussim et al., 1996). Wenn Jungen im Durchschnitt tatsächlich bessere Mathematikleistungen erzielen als Mädchen, würde das Stereotyp „Jungen seien in Mathematik besser als Mädchen“ die Realität akkurat widerspiegeln. Würde eine Lehrkraft dieses Stereotyp nun (bewusst oder unbewusst) bei der Einschätzung der Mädchen und Jungen der Klasse anwenden, dürften daraus auf Gruppenebene relativ akkurate Erwartungen und Einschätzungen resultieren (Jussim et al., 1996; Lorenz, G., 2018). Die Lehrkraft würde den Mädchen dem Stereotyp folgend etwas geringere Fähigkeiten in Mathematik zusprechen, was die tatsächlichen Unterschiede widerspiegeln dürfte. Die Anwendung eines akkuraten Stereotyps kann somit auf Gruppenebene durchaus zu akkuraten Urteilen beitragen. Würde das Stereotyp hingegen einen größeren als tatsächlich existierenden Unterschied zwischen Mädchen und Jungen annehmen oder aussagen, Mädchen seien besser in Mathematik als Jungen, würde die Anwendung dieses inakkuraten Stereotyps zu einer Überschätzung der Unterschiede bzw. zu einer Fehleinschätzung der Richtung des Unterschieds führen. In diesen Fällen trägt die Anwendung des inakkuraten Stereotyps zu inakkurateren Erwartungen und Einschätzungen auf Gruppenebene bei.

Bei Erwartungen und Einschätzungen auf Ebene der individuellen Lernenden dürfte hingegen eine weitgehend eigenschaftsbasierte Verarbeitung stets zu akkurateren Ergebnissen führen als die kategorienbasierte. Dies begründet sich darin, dass die tatsächliche Merkmalsausprägung innerhalb einer sozialen Gruppe variiert. Auch Mädchen untereinander unterscheiden sich in ihrer Mathematikleistung, sodass die individuellen Leistungen um die mittlere Leistung aller Mädchen streuen. Auf individueller Ebene führt demnach selbst die Anwendung eines akkuraten Stereotyps immer dann zu einer Fehleinschätzung, wenn die einzuschätzende Person nicht dem Gruppendurchschnitt entspricht – das einzuschätzende Mädchen also beispielsweise bessere Leistungen in Mathematik erbringt als Mädchen im Durchschnitt.

Auf der Grundlage dieser Überlegungen lässt sich festhalten, dass die Erwartungen und Einschätzungen von Lehrkräften durch verallgemeinernde Überzeugungen über Gruppen von Schüler*innen (Stereotype) beeinflusst sein können. Solche kategorienbasierten Prozesse treten vor allem dann auf, wenn die zeitlichen und kognitiven Ressourcen eingeschränkt sind, geringe Motivation zur Informationssuche vorherrscht sowie wenn nur wenige oder vor allem kategorienkonsistente Informationen zu Verfügung stehen. Die Anwendung stereotyper Annahmen kann dabei auf Gruppenebene durchaus zu akkuraten Erwartungen und Einschätzungen beitragen. Dies ist immer dann der Fall, wenn das Stereotyp tatsächlich existierende Gruppenunterschiede akkurat widerspiegelt. Auf individueller Ebene sollte die Anwendung inakkurater und akkurater Stereotype hingegen oft zu inakkurateren Erwartungen und Einschätzungen führen.

2.2.3 Empirische Befunde zur (In)Akkuratheit von Leistungserwartungen und -einschätzungen

Um das Maß der Akkuratheit bzw. Inakkuratheit von Leistungserwartungen und -einschätzungen von Lehrkräften empirisch zu bestimmen, werden zumeist korrelative Maße herangezogen. Bei diesem Vorgehen wird das Lehrkrafturteil mit einem objektiven Kriterium verglichen. Schätzt beispielsweise eine Lehrkraft die Mathematikleistungen ihrer Schüler*innen ein, werden als Kriterium die Ergebnisse der Lernenden in einem standardisierten Mathematiktest herangezogen. Hohe Korrelationen zwischen beiden Maßen werden als hohe Akkuratheit, geringe Korrelationen als geringe Akkuratheit interpretiert. Dieses Vorgehen ist mit dem Ansatz zur Bestimmung der Rangkomponente der diagnostischen Kompetenz von Lehrkräften vergleichbar, auf den in Kapitel 2.1.3 eingegangen wurde.

Wie in Kapitel 2.1.2 beschrieben, kommen insbesondere bei Leistungserwartungen (Prognosen zukünftiger Leistungen) weitere lernrelevante Merkmale der Schüler*innen als valide Informationsquellen in Betracht. Auch bei diesem breiteren Verständnis von Akkuratheit werden korrelative Maße zur Operationalisierung der (In)Akkuratheit genutzt. Da das Lehrkrafturteil in diesem Fall allerdings nicht mit nur einem Kriterium sondern mit mehreren Kriterien verglichen wird, erfolgt die Bestimmung über multiple Regressionsanalysen. Derjenige Varianzanteil innerhalb der Leistungserwartungen, der von den leistungsrelevanten Merkmalen der Lernenden erklärt wird, bildet die Akkuratheit der Leistungserwartungen ab. Umgekehrt wird diejenige Varianz in den Leistungserwartungen, die nicht durch die genannten Merkmale erklärt werden kann, als Anteil der Inakkuratheit von Leistungserwartungen verstanden (Madon, Jussim & Eccles, 1997).

Empirischen Befunden zufolge, ist ein bedeutsamer Anteil von Leistungserwartungen und -einschätzungen als akkurat zu beurteilen (vgl. Jussim, Robustelli & Cain, 2009). Die geteilte Varianz von Leistungseinschätzungen und mit standardisierten Leistungstests erfassten Leistungen der Lernenden belief sich in Metaanalysen auf etwa 40 % (Hoge & Coladarci, 1989; Südkamp et al., 2012). Hoge und Coladarci (1989) bezogen 16 Studien mit Daten von Lehrkräften und Schüler*innen von Regelschulen in ihre Metaanalyse ein und verglichen die Leistungseinschätzungen von Lehrkräften mit Ergebnissen von Schüler*innen in standardisierten Leistungstests. Die Korrelationen zwischen Lehrkrafturteil und Testergebnis variierten zwischen $r = .28$ und $r = .92$. Der Median der Korrelationen lag bei $r = .66$, was einem geteilten Varianzanteil von etwa 44 % entspricht. Zu einem ähnlichen Ergebnis kommt die Metaanalyse von Südkamp et al. (2012), die auf 75 neueren Studien basiert: Die mittlere Korrelation zwischen Lehrkrafteinschätzung und Testergebnis lag bei $r = .63$, was einem geteilten Varianzanteil von etwa 40 % entspricht.

Wenige Studien haben bislang weitere lernrelevante Merkmale der Schüler*innen in das Maß der Akkuratheit hinzugezogen. Eine Ausnahme ist die aus den Niederlanden stammende

Untersuchung von de Boer et al. (2010), die als Indikator für Leistungserwartungen die Übergangsempfehlung in die Sekundarstufe heranzogen. Auf Seiten der Schüler*innen haben die Autorinnen ein übergreifendes Maß für die aktuellen Leistungen, das sowohl sprachliche und mathematische Kompetenzen als auch metakognitive Fähigkeiten umfasst, ihre allgemeinen kognitiven Fähigkeiten und ihre Leistungsmotivation als valide Basis der Erwartungen berücksichtigt. Die durch diese Variablen erklärte Varianz in der Übergangsempfehlung betrug 67 %.

Gleichzeitig wurde deutlich, dass die Akkuratheit von Lehrkrafturteilen sowohl zwischen Studien als auch innerhalb von Studien zwischen Lehrkräften bedeutsam variiert (z. B. Hosenfeld, Helmke & Schrader, 2002; Karing, Matthäi & Artelt, 2011; Südkamp et al., 2012). Südkamp et al. (2012) identifizierten vier Merkmalsbereiche die die Urteilsakkuratheit beeinflussen können (Modell der Genauigkeit von Lehrkräfteurteilen): Merkmale (1) des Urteils selbst, (2) des Kriteriums mit dem ein Urteil verglichen wird, (3) der Lehrkraft und (4) der Schüler*innen. Diese Systematisierung soll herangezogen werden, um den empirischen Forschungsstand zu Bedingungsfaktoren für die Akkuratheit von Lehrkrafturteilen zu beschreiben (vgl. dazu auch Hoffmann, 2017).

Merkmale des Urteils

Im Zusammenhang mit den Merkmalen des Urteils werden primär drei Aspekte diskutiert. Die Urteile können sich darin unterscheiden, wie genau eine Lehrkraft über das Maß – beispielsweise einen Leistungstest – informiert ist, mit dem ihr Urteil später verglichen wird (Informiertheit). Außerdem variieren die Urteile bezüglich der Spezifität ihrer Skalen (Urteilsspezifität) und darin, ob das Urteil eher einen globalen Eindruck abbildet oder sich auf sehr spezifische (Teil-)Kompetenzen bezieht (Domänenspezifität). Der angenommene positive Zusammenhang zwischen *Informiertheit* und Urteilsakkurateit wurde von bisherigen Studien weitgehend bestätigt und verdeutlicht, dass Lehrkräfte tendenziell zu genaueren Urteilen gelangen, wenn sie den Vergleichsmaßstab kennen (Demaray & Elliott, 1998; Feinberg & Shapiro, 2003, 2009; Hoge & Coladarci, 1989; Südkamp et al., 2012). Die Befundlage zur *Urteilsspezifität* ist gemischt. Während die Metaanalyse von Hoge und Coladarci (1989) andeutete, dass Urteile, die anhand einer Ratingskala (geringe Spezifität) vorgenommen wurden, weniger akkurat ausfielen als Urteile anhand von Rangreihen, notenäquivalenten Skalen, Anzahl korrekter Antworten oder itemspezifischen Einschätzungen (aufsteigend höhere Spezifität), fanden Südkamp et al. (2012), ebenfalls in einer Metaanalyse, keinen Hinweis auf Unterschiede in Abhängigkeit von der Spezifität der verwendeten Skala. Gleiches gilt für den Grad der *Domänenspezifität*. Bisherige Studien finden keinen Hinweis auf Unterschiede zwischen globaleren Einschätzungen und Urteilen über sehr spezifische Teilkompetenzen per se (Südkamp et al., 2012). Stattdessen scheinen Urteile akkurater auszufallen, wenn das Urteil und das

herangezogene Kriterium in ihrer Domänenspezifität kongruieren – also sich beispielsweise beide auf die wortbasierte Leseflüssigkeit beziehen (Begeny, Krouse, Brown & Mann, 2011; Südkamp et al., 2012).

Zudem wird teilweise angenommen, dass sich die Akkuratheit von Lehrkräfteurteilen zwischen verschiedenen Schulfächern unterscheidet. Wenngleich die Befundlage hierzu nicht ganz einheitlich ist, scheinen keine systematischen Unterschiede zwischen der Akkuratheit von Urteilen im sprachlichen und mathematischen Bereich zu bestehen (Demaray & Elliott, 1998; Hinnant et al., 2009; Südkamp et al., 2012). Gleichzeitig lassen sich Hinweise darauf finden, dass es Lehrkräften, die sowohl Deutsch als auch Mathematik unterrichten, oft in einem der beiden Fächer besser gelingt, zu akkuraten Urteilen zu gelangen, als im jeweils anderen Fach (Lorenz, C. & Artelt, 2009).

Merkmale des Kriteriums

Auf Seiten des Kriteriums werden insbesondere zwei Aspekte diskutiert. Zum einen wird angenommen, dass es Lehrkräften leichter fallen sollte, zu akkuraten Urteilen zu gelangen, wenn sie sich auf sogenannte *curriculum based measurements (CBM)* beziehen (z. B. Feinberg & Shapiro, 2003; Madelaine & Wheldall, 2005). CBM-Verfahren sind eng am schulischen Curriculum ausgerichtet und dienen dazu, Lehrkräften Rückmeldungen über die Leistungen ihrer Schüler*innen zu geben. Bisherige Forschung unterstützt diese Annahme nicht. So fanden Südkamp et al. (2012) in ihrer Metaanalyse keine Unterschiede in der Urteilsakkuratheit, die sich auf CBM-Verfahren als Kriterium bezog verglichen mit typischen Leistungstests als Kriterium.

Zudem wird teilweise davon ausgegangen, dass die Urteilsgüte vom zeitlichen Abstand zwischen dem Urteil der Lehrkraft und der Erfassung des Kriteriums abhängt. Die Annahme, dass eine stärkere Übereinstimmung erzielt wird, wenn das Urteil zeitlich näher zur Erfassung des Kriteriums liegt, konnte metaanalytisch nicht bestätigt werden (Südkamp et al., 2012).

Merkmale der Lehrkraft

Aufgrund der insgesamt hohen Variation der Urteilsgüte zwischen Lehrkräften (z. B. Anders, Kunter, Brunner, Krauss & Baumert, 2010; Hosenfeld et al., 2002; Karing et al., 2011; Lorenz, C. & Artelt, 2009) liegt es nahe, dass die Urteilsgüte von Merkmalen der Lehrkräfte mitbestimmt sein könnte. Dabei wird zunächst angenommen, dass Lehrkräfte mit längerer Berufserfahrung mehr Übung in der Einschätzung von Schüler*innen haben und daher zu akkurateren Urteilen gelangen sollten (vgl. z. B. Coladarci, 1986; Impara & Plake, 1998). Empirische Studien, die sich mit dieser Frage auseinandergesetzt haben, finden jedoch nur sehr kleine, zumeist nicht signifikante Zusammenhänge

(z. B. Bates & Nettelbeck, 2001; Feinberg & Shapiro, 2003; Wild & Rost, 1995; vgl. auch Anders, Kunter, et al., 2010; Oerke, McElvany, Ohle, Ullrich & Horz, 2016). Lediglich in der Studie von McElvany et al. (2009) zeigten Lehrkräfte mit längerer Berufserfahrung eine höhere diagnostische Sensitivität bei der Einschätzung individueller Schüler*innenleistungen.

Alternativ wird vermutet, dass vielleicht nicht Berufserfahrung per se, sondern spezifischer die fachdidaktischen Kompetenzen der Lehrkräfte zu einer höheren Urteilsakkuratheit beitragen (vgl. z. B. McElvany et al., 2009). Jedoch wird auch diese Annahme durch Ergebnisse bisheriger Forschung nur bedingt unterstützt (McElvany et al., 2009). Stattdessen weisen einzelne Studien auf qualitative Unterschiede im Urteilsprozess zwischen Lehramtsanwärterinnen bzw. -anwärtern und erfahrenen Lehrkräften hin. Erfahrenen Lehrkräften scheint es besser zu gelingen, die Tiefe und Intensität der Informationsverarbeitung an der Bedeutsamkeit des zu treffenden Urteils auszurichten (z. B. Krolak-Schwerdt, Böhmer & Gräsel, 2009, 2012; Krolak-Schwerdt & Rummer, 2005) und ihr Urteil bei inkonsistenten Informationen flexibler an die neuen Informationen anzupassen (z. B. Ophuysen, 2006).

Über diese spezifisch professionellen Merkmale der Lehrkräfte hinaus wird zudem angenommen, dass die Urteilsgüte umso höher ausfallen sollte, je länger sich Lehrkraft und Schüler*innen kennen, da die Lehrkraft dann auf mehr Informationen über die einzelnen Schüler*innen zurückgreifen kann. Entgegen dieser Annahme zeigten sich nur geringe, nicht signifikante Korrelationen zwischen der Urteilsakkuratheit und der Kontaktdauer (z. B. Hoffmann & Böhme, 2014; Lorenz, C., 2011; Oerke et al., 2016; Schrader, 1989; Wild & Rost, 1995).

Merkmale der Schüler*innen

Verschiedene Studien unterstützen, dass es Lehrkräften häufig schwerfällt, Schüler*innen akkurat zu identifizieren, die besonders hohe (z. B. Eckert, Dunn, Coddington, Begeny & Kleinmann, 2006) oder besonders niedrige (z. B. Artelt, Stanat, Schneider & Schiefele, 2001; Begeny, Eckert, Montarello & Storie, 2008; Begeny et al., 2011; Madelaine & Wheldall, 2005) Leistungen zeigen. Gleichzeitig scheint es Lehrkräften leichter zu fallen, die Leistungen ihrer Schüler*innen in akkurate Rangreihen zu bringen (Rangkomponente), wenn die Leistungen in einer Klasse heterogener sind (z. B. Rjosk, McElvany, Anders & Becker, 2011; Schrader, 1989; Weinert, Schrader & Helmke, 1990). Referenzgruppeneffekte, denen zufolge identische individuelle Leistungen in Klassen mit niedrigerem mittleren Leistungsniveau höher eingeschätzt werden als in Klassen mit höherem mittlerem Leistungsniveau, wurden hingegen nur bei unspezifischen Urteilen und bei einer geringen Informiertheit der Lehrkräfte identifiziert (Südkamp & Möller, 2009). Die Anzahl an Schüler*innen in

der Klasse scheint ebenfalls von geringer Bedeutung zu sein (z. B. Anders, Kunter, et al., 2010; Wild & Rost, 1995).

Ein anderer Bereich von Einflüssen auf die Urteilsakkuratheit bezieht sich auf weitere leistungsnahe Merkmale der Schüler*innen. Das Wissen oder eine Annahme über solche Merkmale können Leistungseinschätzungen im Sinne von Halo-Effekten (Überstrahleffekt; Gesamteindruck einer Person bestimmt die Wahrnehmung einzelner Merkmale) oder durch logische Fehler (Einfluss einer impliziten Persönlichkeitstheorie; man schließt von der Leistung in einem Fach bzw. Bereich auf die Leistung in einem anderen Fach bzw. Bereich) beeinflussen. So wurden unter anderem das Wissen zu oder die Annahme über die Leistung in einem anderen Fach (Dompnier, Pansu & Bressoux, 2006), eine Klassenwiederholung oder die wahrgenommene Lernmotivation (Dompnier et al., 2006; Kaiser, Retelsdorf, Südkamp & Möller, 2013; Urhahne et al., 2010), die allgemeinen kognitiven Fähigkeiten (z. B. Hoge & Butcher, 1984; Schrader & Helmke, 1990) und das Lernverhalten (Bennett, Gottesman, Rock & Cerullo, 1993) als einflussnehmende Faktoren identifiziert.

Gleichzeitig wurden verzerrende Effekte leistungsferner Merkmale der Schüler*innen beobachtet. Hier spielen vor allem Merkmale wie ein Zuwanderungshintergrund, der sozioökonomische Status der Familien der Schüler*innen sowie ihr Geschlecht eine Rolle. Ergebnisse zu diesen Merkmalen werden im folgenden Kapitel genauer beschrieben. Aber auch weitere leistungsferne Merkmale wie Mehrsprachigkeit (z. B. Rjosk et al., 2011), Verhaltensprobleme (z. B. Bennett et al., 1993; Hecht & Greenfield, 2002) oder das Vorliegen eines sonderpädagogischen Förderbedarfs (z. B. Hurwitz, Elliott & Braden, 2007; Jenkins & Demaray, 2016; Shifrer, 2013, 2016) können Leistungserwartungen und -einschätzungen von Lehrkräften beeinflussen.

2.2.4 Empirische Befunde zu Verzerrungen von Leistungserwartungen und -einschätzungen

Im Folgenden wird der empirische Forschungsstand dazu zusammengefasst, ob und inwiefern Verzerrungen in Leistungserwartungen und -einschätzungen von Lehrkräften in Abhängigkeit des ethnischen Hintergrunds, sozialen Hintergrund oder Geschlechts der Lernenden verzerrt sind. Dabei werden auch Forschungsdesiderate herausgearbeitet.

Verzerrungen in Abhängigkeit vom Zuwanderungshintergrund

Verschiedene internationale Studien weisen darauf hin, dass Lehrkräfte häufig Kindern, die ethnischen Minderheiten angehören oder aus zugewanderten Familien stammen, geringere Leistungen zusprechen, sodass ihre Erwartungen und Einschätzungen für diese Kinder geringer ausfallen als für Kinder ohne Zuwanderungshintergrund (Dusek & Joseph, 1983; Jussim et al., 1996;

Ready & Wright, 2011; Riegle-Crumb & Humphries, 2012; Tenenbaum & Ruck, 2007; van den Bergh, Denessen, Hornstra, Voeten & Holland, 2010). In vielen Fällen lassen sich diese niedrigeren Leistungserwartungen und -einschätzungen vollständig auf tatsächliche Leistungsunterschiede zwischen den Gruppen zurückführen (Alvidrez & Weinstein, 1999; de Boer et al., 2010; Hartley, 1982; Hinnant et al., 2009; Jussim et al., 1996; Madon et al., 1998; Riegle-Crumb & Humphries, 2012; Rubie-Davies & Peterson, 2016; Van Houtte, Demanet & Stevens, 2013). Teilweise weisen empirische Befunde allerdings auch darauf hin, dass die Leistungserwartungen und -einschätzungen für Lernende mit Zuwanderungshintergrund ungerechtfertigt niedrig ausfallen und in Abhängigkeit des Zuwanderungshintergrunds negativ verzerrt sind (McKown & Weinstein, 2008; Meissel et al., 2017; Ready & Wright, 2011; Timmermans, de Boer & van der Werf, 2016). So untersuchten Ready und Wright (2011) beispielsweise auf der Basis von knapp 9500 Kindergartenkindern und ihren Fachkräften in den USA, ob die Einschätzungen der Literacy-Kompetenzen in Abhängigkeit des ethnischen Hintergrunds der Kinder verzerrt sind. Die Ergebnisse bestätigen diese Annahme. Auch nach Kontrolle der individuellen Ergebnisse in einem Literacy-Test wurden die Literacy-Kompetenzen von Kindern mit Zuwanderungshintergrund geringer eingeschätzt als von Kindern ohne Zuwanderungshintergrund.

In Deutschland liegen bislang kaum Ergebnisse zu Verzerrungen in Leistungserwartungen und -einschätzungen von Lehrkräften in Abhängigkeit vom Zuwanderungshintergrund der Lernenden vor. Die nur vereinzelt vorliegenden Studien zur diagnostischen Kompetenz von Lehrkräften fanden nach Kontrolle standardisierter Leistungsmaße keine Unterschiede in den Leistungseinschätzungen für Kinder mit und Kinder ohne Zuwanderungshintergrund (Hachfeld, Anders, Schroeder, Stanat & Kunter, 2010; Rjosk et al., 2011; Schrader & Helmke, 1990; Stahl, 2007). Diese Untersuchungen differenzierten jedoch nicht zwischen Schüler*innen verschiedener Herkunftsländer und lassen damit keinen Rückschluss auf eventuelle Unterschiede zwischen verschiedenen ethnischen Herkunftsgruppen zu. Da sich Stereotype zumeist auf spezifische Herkunftsgruppen beziehen und sich entsprechend auch die darauf basierenden Annahmen für verschiedene Gruppen unterscheiden dürften, sind differenziertere Analysen jedoch notwendig. Vier neuere Vignetten-Studien in Deutschland haben spezifisch Leistungseinschätzungen für Schüler*innen mit türkischem Zuwanderungshintergrund untersucht. Die Studien kommen auf Basis von Vignetten von Schülern, die entweder keinen oder einen türkischen Zuwanderungshintergrund hatten, zu unterschiedlichen Ergebnissen. Während die Ergebnisse von Glock und Krolak-Schwerdt (2014) zeigten, dass Lehrkräfte unter anderem die sprachlichen und allgemeinen kognitiven Fähigkeiten sowie die Leistungen in den Unterrichtsfächern Deutsch und Mathematik für beide Vignetten gleich einschätzten, fand die Studie von Tobisch und Dresel (2017) für alle untersuchten Lehrkrafturteilen signifikante Unterschiede in Abhängigkeit des Zuwanderungshintergrunds. So schätzten die Lehrkräfte unter anderem die

allgemeinen kognitiven Fähigkeiten sowie die erwarteten Noten in Deutsch und Mathematik für die Vignette eines Jungen mit türkischem Zuwanderungshintergrund geringer ein als für die Vignette eines Jungen ohne Zuwanderungshintergrund. Auch in der Studie von Holder und Kessels (2017) wurden die objektiven Lese- und Schreibfähigkeiten von Kindern mit türkischem Zuwanderungshintergrund eher unterschätzt. Die Studie von Glock und Krolak-Schwerdt (2013) fand systematische Unterschätzungen der Lese-, Schreib- und Sprachleistungen für Schüler mit türkischem Zuwanderungshintergrund, wenn Vignetten von leistungsschwachen Lernenden untersucht wurden, aber keine Unterschiede in den Leistungseinschätzungen bei leistungsstarken Lernenden.

Verzerrungen in Abhängigkeit vom sozioökonomischen Hintergrund

Auch zum sozioökonomischen Hintergrund der Schüler*innen liegen international Hinweise für Verzerrungen in Leistungserwartungen und -einschätzungen vor. So werden Schüler*innen aus Familien mit einem niedrigeren sozialen Status auch nach Kontrolle der gemessenen Leistungen häufig als leistungsschwächer eingeschätzt als Kinder aus sozial bessergestellten Familien (Alvidrez & Weinstein, 1999; de Boer et al., 2010; Jussim & Eccles, 1995; Jussim et al., 1996; Madon et al., 1998; Ready & Wright, 2011; Timmermans et al., 2016; Timmermans, Kuyper & van der Werf, 2015; Van Houtte et al., 2013; van Matre, Valentine & Cooper, 2000).

Die im deutschsprachigen Raum vorliegenden Befunde zu sozialen Verzerrungen in Leistungserwartungen und -einschätzungen von Lehrkräften sind hingegen uneinheitlich. Einige Feldstudien fanden keine Verzerrungen in Abhängigkeit des sozialen Status (Hachfeld et al., 2010; Karing et al., 2011). Stahl (2007) berichtete hingegen häufigere Überschätzungen im Bereich Lesen für Kinder aus sozial bessergestellten Familien. Eine weitere Feldstudie aus Deutschland untersuchte Erwartungen von Lehrkräften für die Wahrscheinlichkeit, das Abitur zu erreichen (Becker, 2013). Bei Jugendlichen der zehnten Jahrgangsstufe überschätzten die Lehrkräfte diese Wahrscheinlichkeit für Schüler*innen aus sozial bessergestellten Familien auch nach Kontrolle ihrer Leistungen. Auch neuere Vignetten-Studien in Deutschland kommen zu unterschiedlichen Ergebnissen. Während die Leistungseinschätzungen in der Studie von Glock und Krolak-Schwerdt (2014) nicht in Abhängigkeit des sozialen Status variierten, fanden Tobisch und Dresel (2017) in allen untersuchten Lehrkrafturteilen – außer den erwarteten Noten im Fach Mathematik – signifikante Unterschiede zugunsten der Vignette eines Kindes mit höherem sozioökonomischen Status.

Verzerrungen in Abhängigkeit vom Geschlecht

Die geschlechtsspezifischen Befunde weisen auf bereichsspezifische Effekte hin. So wurden internationalen Befunden zufolge die Leistungen von Mädchen auch nach Kontrolle von Testergebnissen vor allem im sprachlichen Bereich (Hartley, 1982; Hinnant et al., 2009; Meissel et al., 2017; Ready & Wright, 2011) und in bereichsübergreifenden Leistungsmaßen (de Boer et al., 2010; Hartley, 1982; Ross & Jackson, 1991; Timmermans et al., 2016; Timmermans et al., 2015; Van Houtte et al., 2013; van Matre et al., 2000; Wood, Kaplan & McLoyd, 2007) höher eingeschätzt als die der gleich leistungsstarken Jungen. In Mathematik hingegen werden häufig die Kompetenzen der Mädchen zu niedrig eingeschätzt (Jussim & Eccles, 1992; Riegle-Crumb & Humphries, 2012; Rubie-Davies & Peterson, 2016). Ähnlich viele Studien identifizierten jedoch keine Verzerrungen in Abhängigkeit vom Geschlecht der Lernenden im sprachlichen Bereich (Doherty & Conolly, 1985; Hoge & Butcher, 1984; McKown & Weinstein, 2008), in Mathematik (Doherty & Conolly, 1985; Hinnant et al., 2009; Jussim, 1989; Jussim & Eccles, 1995; Jussim et al., 1996; Madon et al., 1998; McKown & Weinstein, 2008) und für bereichsübergreifende Leistungserwartungen und -einschätzungen (Alvidrez & Weinstein, 1999; Hoge & Coladarci, 1989).

Auch im deutschen Kontext sind die Studienergebnisse zu geschlechtsbezogenen Verzerrungen in Leistungserwartungen und -einschätzungen uneinheitlich. Einige Studien berichteten keine geschlechtsbezogenen Unterschiede in den Leistungseinschätzungen von Lehrkräften, wenn die Leistungen der Kinder kontrolliert wurden (Lesen: Karing et al., 2011; Mathematik: Schrader & Helmke, 1990). Stahl (2007) berichtete hingegen Überschätzungen für Mädchen im Lesen und Tiedemann (2000, 2002) fand negativ verzerrte Einschätzungen für Mädchen in Mathematik.

Zusammenfassung

Zwar deutet eine Vielzahl der Studien darauf hin, dass Erwartungen und Einschätzungen von Lehrkräften für die Leistungen von Lernenden in Abhängigkeit von deren soziodemografischen Merkmalen verzerrt sein können. Die empirische Befundlage ist allerdings gemischt. Zudem liegen für den deutschsprachigen Bildungskontext bislang nur sehr wenige Studien vor. Weitere Forschungsanstrengungen zur Frage, ob Leistungserwartungen und -einschätzungen von Lehrkräften verzerrt sind, erscheinen daher sinnvoll. Dabei sollten mit Blick auf den Zuwanderungshintergrund verschiedene Herkunftsgruppen miteinander verglichen werden. Gleichmaßen ist zu berücksichtigen, dass der ethnische und soziale Hintergrund von Schüler*innen konfundiert ist, was sich auch im Auftreten von Verzerrungen in Leistungserwartungen und -einschätzungen niederschlagen kann (Jussim et al., 1996). Hinsichtlich etwaiger Einflüsse vom Geschlecht der

Schüler*innen ist die empirische Befundlage besonders uneinheitlich. Zentral scheint hier zu sein, in zukünftigen Untersuchungen bereichsspezifische Analysen durchzuführen.

2.2.5 Empirische Befunde zu Einschätzungen der Motivation und des Arbeitsverhaltens

In diesem Kapitel sollen Lehrkräfteeinschätzungen zur Motivation und zum Arbeitsverhalten von Schüler*innen fokussiert werden. Einschätzungen dieser und verwandter Merkmale stehen in einem engen Zusammenhang mit Leistungserwartungen und -einschätzungen, weshalb ihnen bei der Untersuchung von Verzerrungen in Leistungserwartungen besondere Bedeutung zukommt. Verschiedene Studien zeigen, dass Lernende, denen die Lehrkraft eine hohe Motivation und ein günstiges Arbeitsverhalten zuspricht, auch als leistungsstärker eingeschätzt werden (Anders, McElvany & Baumert, 2010; Fitzpatrick, Côté-Lussier & Blair, 2016; Kaiser et al., 2013; Kuhl & Hannover, 2012; Timmermans et al., 2016; Trautwein & Baeriswyl, 2007). So stand in der Studie von Anders, McElvany, et al. (2010) die Urteilsdimension „Motivation und Lerntugenden“, die Lehrkräfteeinschätzungen zur Aufmerksamkeit im Unterricht, zum Interesse am Lernen, zur Lern- und Leistungsmotivation, Anstrengungsbereitschaft, Ausdauer, Gewissenhaftigkeit, Pünktlichkeit und Disziplin sowie zum Fleiß und Ehrgeiz der Schüler*innen enthält, in einem positiven mittleren Zusammenhang mit den (rekodierten) Noten und der Übergangsempfehlung ($r = .60$ bis $.67$). Auch in der Studie von Trautwein und Baeriswyl (2007) war die Lehrkräfteeinschätzung zur Motivation und zum Arbeitsverhalten positiv mit der gemittelten Einschätzung der Lernstände in Mathematik und Deutsch ($r = .49$) sowie der Übergangsempfehlung ($r = .50$) assoziiert. Ebenso stand die Lehrkräfteeinschätzung des Engagements der Schüler*innen im Lesen in engem Zusammenhang mit der Lehrkräfteeinschätzung der Leseleistung ($r = .79$; Kaiser et al., 2013). Dass diese Assoziationen nicht ausschließlich auf tatsächlich bestehende Leistungsunterschiede zwischen den Schüler*innen zurückgehen, zeigen aktuelle Studien von Robinson-Cimpian et al. (2014) und Timmermans et al. (2016). Auch nach Kontrolle umfangreicher Leistungsmaße und Hintergrundmerkmale sagte die Lehrkräfteeinschätzung des Arbeitsverhaltens der Lernenden die Leistungserwartungen bzw. -einschätzungen der Lehrkräfte signifikant positiv vorher.

Diese Zusammenhänge verdeutlichen, dass Lehrkräfte ihre Leistungserwartungen und -einschätzungen nicht ausschließlich auf verfügbaren Leistungsinformationen gründen, sondern zusätzlich ihre Einschätzung weiterer lernrelevanter Merkmale der Lernenden, wie etwa der Motivation oder des Arbeitsverhaltens, in ihre Urteile einbeziehen. Zu der Frage, zu welchem Grad diese Lehrkräfteeinschätzungen akkurat sind, liegen bislang nur relativ wenige Studien vor. Diese Studien weisen darauf hin, dass Lehrkräfte die intrinsische Lernmotivation (Spinath, 2005), das

schulische Interesse (Karing, 2009; Rellensmann & Schukajlow, 2017) sowie die Lernzielorientierung (Dicke, Lüdtke, Trautwein, Nagy & Nagy, 2012; Givvin, Stipek, Salmon & MacGyvers, 2001; Urhahne, Chao, Florineth, Luttenberger & Paechter, 2011), die Bereitschaft zur Beteiligung an lesebezogenen Unterrichtsaktivitäten (Kaiser et al., 2013), das akademische Selbstkonzept (Givvin et al., 2001; Praetorius, Berner, Zeinz, Scheunpflug & Dresel, 2013; Spinath, 2005; Urhahne et al., 2011) und positive bzw. negative Emotionen beim Lernen bzw. die Leistungsängstlichkeit (Givvin et al., 2001; Spinath, 2005; Urhahne et al., 2011) nur relativ ungenau einschätzen (für einen aktuellen Überblick siehe Praetorius & Südkamp, 2019). Als Vergleichsmaß für die Lehrkrachteinschätzungen werden dabei typischerweise Selbsteinschätzungen der Schüler*innen herangezogen. Dieses Vorgehen basiert auf der Annahme, dass die betreffenden Personen selbst die verlässlichsten Aussagen über ihre Motivation, Emotionen und Selbstkognitionen, die letztlich innerliche Zustände darstellen, treffen können (vgl. z. B. Givvin et al., 2001). Spinath (2005) verglich beispielsweise die Lehrkrachteinschätzungen von 43 Klassenlehrkräften der ersten bis vierten Klasse mit den Selbsteinschätzungen ihrer 723 Schüler*innen. Untersucht wurden die Übereinstimmungen der Lehrkraft- und Selbsteinschätzungen hinsichtlich des akademischen Selbstkonzepts, der intrinsischen Lernmotivation und der Leistungsängstlichkeit. Während die Korrelation für das akademische Selbstkonzept mit $r = .39$ im mittleren Bereich lag, zeigten sich für die Lernmotivation ($r = .20$) und Leistungsängstlichkeit ($r = .15$) nur niedrige Korrelationen zwischen der Lehrkrachteinschätzung und der Selbsteinschätzung der Lernenden. Verglichen mit den für Leistungsmaße ermittelten Korrelationen von $r = .60$ bis $.70$ (Südkamp et al., 2012) sind diese Übereinstimmungen als gering einzustufen.

Diese geringen Übereinstimmungen können einerseits durch zufällige Abweichungen zustande kommen. Andererseits können auch für die Lehrkrachteinschätzungen der Motivation und des Arbeitsverhaltens weitere Merkmale der Schüler*innen relevant sein. Naheliegender wäre es, dass auch bei der Bildung dieser Urteile soziodemografische Merkmale, wie etwa der ethnische Hintergrund, soziale Hintergrund oder das Geschlecht der Lernenden, eine Rolle spielen. Ob Lehrkrachteinschätzungen der Motivation und des Arbeitsverhaltens für diese Gruppen von Lernenden bedeutsam variieren, wurde bislang in nur wenigen Studien untersucht. Die Befunde sind zudem uneinheitlich und beziehen nur sehr selten Selbsteinschätzungen der Schüler*innen oder objektive Indikatoren der Motivation und des Arbeitsverhaltens ein. Die Analysen liefern daher nur selten Erkenntnisse dazu, inwieweit etwaige Gruppenunterschiede in den Lehrkrachteinschätzungen Unterschieden in den Selbsteinschätzungen der Schüler*innen entsprechen oder ob sie eher auf stereotypen Annahmen über diese Merkmale basieren.

Im Hinblick auf den *ethnischen Hintergrund* von Schüler*innen zeigte sich in einer Vignetten-Studie aus den USA, dass Grundschullehrkräfte Jungen mit kaukasisch-amerikanisch klingendem Namen eine höhere schulische Motivation zusprachen als Jungen mit afro-amerikanisch klingendem

Namen (Anderson-Clark, Green & Henley, 2008). In einer ebenfalls in den USA durchgeführten Feldstudie wurde hingegen kein Unterschied in der Lehrkräfteeinschätzung der Anstrengungsbereitschaft für Schüler*innen mit und Schüler*innen ohne Zuwanderungshintergrund identifiziert, was mit den Selbsteinschätzungen der Lernenden übereinstimmte (Jussim et al., 1996). Im Gegensatz dazu schätzten Lehrkräfte in einer belgischen Studie Kinder mit Zuwanderungshintergrund auch nach Kontrolle ihrer Selbsteinschätzungen weniger anstrengungsbereit im Unterricht ein als Kinder ohne Zuwanderungshintergrund. Die Lehrkräfteeinschätzung dazu, wie fleißig die Lernenden ihre Hausaufgaben bearbeiten, variierte hingegen nicht in Abhängigkeit des Zuwanderungshintergrunds (Van Houtte et al., 2013).

In Deutschland wurden Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens für Kinder mit und Kinder ohne Zuwanderungshintergrund bislang in zwei Feldstudien verglichen. Diesen Studien zufolge nehmen Lehrkräfte für Kinder ohne Zuwanderungshintergrund eine etwas höhere Motivation und ein etwas günstigeres Arbeitsverhalten (Anders, McElvany, et al., 2010) sowie eine leicht höhere Schulfreude, Anstrengungsbereitschaft und günstigere schulische Arbeitshaltung an (Schneider, 2011) als für Kinder aus zugewanderten Familien. Selbsteinschätzungen der Lernenden wurden allerdings nicht kontrolliert. Ebenfalls wurden keine ethnischen Gruppen in den Analysen unterschieden. Zwei neuere experimentelle Studien in Deutschland untersuchten Lehrkräfteeinschätzungen motivationaler Merkmale und des Arbeitsverhaltens für Kinder mit türkischem Zuwanderungshintergrund im Vergleich zu Kindern ohne Zuwanderungshintergrund. In der Studie von Glock und Krolak-Schwerdt (2014) schätzten die Lehrkräfte das schulische Arbeitsverhalten für Kinder mit türkischem und für Kinder ohne Zuwanderungshintergrund gleich ein. In der Vignetten-Studie von Tobisch und Dresel (2017) beschrieben die Lehrkräfte Lernende mit türkischem Zuwanderungshintergrund hingegen als weniger anstrengungsbereit als Lernende ohne Zuwanderungshintergrund.

Auch bezüglich des *sozialen Hintergrunds* der Lernenden haben einzelne Studien Unterschiede in Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens ihrer Schüler*innen untersucht. In einer Feldstudie in den USA schätzten die Lehrkräfte Kinder aus sozial bessergestellten Familien als anstrengungsbereiter ein als Kinder aus sozial benachteiligten Familien, wobei dieser von den Lehrkräften angenommene Unterschied größer ausfiel als in den Selbsteinschätzungen der Lernenden (Jussim et al., 1996). Im Gegensatz dazu unterschieden sich die Anstrengungsbereitschaft im Unterricht und die Sorgfalt beim Erledigen von Hausaufgaben aus Sicht von Lehrkräften in Belgien nicht zwischen Lernenden aus sozial bessergestellten und Lernenden aus sozial schlechtergestellten Familien (Van Houtte et al., 2013). Ebenso fand eine kanadische Studie keinen Unterschied in der Lehrkräfteeinschätzung des Unterrichtsverhaltens von Viertklässler*innen in

Abhängigkeit des sozialen Hintergrunds der Kinder, wenn weitere Hintergrundmerkmale und das Verhalten der Lernenden im Kindergarten kontrolliert wurden (Fitzpatrick et al., 2016).

In Deutschland weisen zwei Feldstudien darauf hin, dass Kinder aus sozial bessergestellten Familien als motivierter von ihren Lehrkräften eingeschätzt werden als Kinder aus sozial benachteiligten Familien (Anders, McElvany, et al., 2010; Trautwein & Baeriswyl, 2007). Dieser Unterschied konnte in der Studie von Trautwein und Baeriswyl (2007) auf Leistungsunterschiede der Lernenden zurückgeführt werden. Zudem kommen zwei neuere Vignetten-Studien zu unterschiedlichen Ergebnissen. Glock und Krolak-Schwerdt (2014) identifizierten in ihrer Studie keinen Unterschied in der Lehrkräfteeinschätzung des schulischen Arbeitsverhaltens in Abhängigkeit des sozialen Hintergrunds der Schüler*innen, während die Lehrkräfte in der Studie von Tobisch und Dresel (2017) Kinder aus sozial bessergestellten Familien als anstrengungsbereiter einschätzten als Kinder aus sozial benachteiligten Familien.

In Abhängigkeit vom *Geschlecht der Lernenden* schätzen Lehrkräfte mehreren nationalen und internationalen Studien zufolge die Motivation und das Arbeitsverhalten der Schüler*innen unterschiedlich ein. Insgesamt weisen die Befunde darauf hin, dass Lehrkräfte annehmen, Mädchen seien im Vergleich zu Jungen motivierter (Anders, McElvany, et al., 2010; Schneider, 2011; Trautwein & Baeriswyl, 2007), anstrengungsbereiter (Jussim & Eccles, 1992; Jussim et al., 1996; Van Houtte et al., 2013), würden ihre Hausaufgaben sorgfältiger erledigen (Van Houtte et al., 2013), sich im Unterricht angemessener verhalten (Fitzpatrick et al., 2016) und über eine höhere schulische Selbstregulationskompetenz verfügen (Duckworth & Seligman, 2006; Kuhl & Hannover, 2012). Inwieweit diese angenommen Geschlechterunterschiede mit Selbsteinschätzungen oder objektiven Maßen kongruieren, bleibt bislang uneindeutig. In der Studie von Duckworth und Seligman (2006) stimmten die Geschlechterunterschiede in der Selbstregulationskompetenz zwischen den Lehrkräfteeinschätzungen und den Selbsteinschätzungen der Lernenden überein. In anderen Studien blieben die Geschlechterunterschiede in den Lehrkräfteeinschätzungen jedoch auch nach Kontrolle von Selbsteinschätzungen bestehen. So differierten in den Studien von Jussim und Eccles (1992) und Jussim et al. (1996) die Einschätzungen der Lehrkräfte und die Angaben der Mädchen und Jungen dazu deutlich, wie sehr sie sich in der Schule anstrengen und wie viel Zeit sie in Hausaufgaben investieren. Ebenso wurden Mädchen auch dann noch als anstrengungsbereiter im Unterricht und sorgfältiger bei den Hausaufgaben eingeschätzt, wenn ihr selbsteingeschätztes Engagement, schulisches Wohlbefinden und Fehlverhalten kontrolliert wurden (Van Houtte et al., 2013). Auch in der Studie von Fitzpatrick et al. (2016) schätzten Lehrkräfte das Unterrichtsverhalten von Mädchen in der vierten Klasse auch nach Kontrolle ihres Lernverhaltens im Kindergarten günstiger ein.

Zusammenfassend weisen die Befunde darauf hin, dass Erwartungen und Einschätzungen von Lehrkräften zur Leistung von Schüler*innen eng damit verbunden sind, wie die Lehrkräfte die

Motivation und das Arbeitsverhalten der Lernenden einschätzen. Gleichzeitig wurde deutlich, dass Lehrkräfte die Motivation und das Arbeitsverhalten verglichen mit Leistungen offenbar nur relativ ungenau einschätzen können. Ebenfalls deuten die vorliegenden Befunde darauf hin, dass Lehrkräfte die Motivation und das Arbeitsverhalten von Lernenden in Abhängigkeit ihrer soziodemografischen Merkmale unterschiedlich beurteilen. Inwieweit diese Unterschiede in den Lehrkräfteeinschätzungen mit Selbsteinschätzungen oder objektiven Maßen kongruieren, bleibt dabei bislang uneindeutig.

2.3 Erwartungen von Lehrkräften und die Leistungsentwicklung von Schüler*innen

In den vorangegangenen Kapiteln wurden Erwartungen und Einschätzungen von Lehrkräften konzeptionell eingebettet und die Urteilsbildung näher beleuchtet. Es wurde deutlich, dass Erwartungen und Einschätzungen von Lehrkräften darin variieren können, wie genau sie die eingeschätzten Merkmale der Lernenden widerspiegeln. Kategorisierende Urteilsprozesse können darüber hinaus dazu beitragen, dass Lehrkrafturteile von soziodemografischen Merkmalen der Schüler*innen, wie etwa von ihrem ethnischen Hintergrund, sozialen Hintergrund oder Geschlecht, beeinflusst werden.

In diesem Kapitel geht es nun um die Bedeutung von Leistungserwartungen und -einschätzungen für die Leistungsentwicklung von Schüler*innen. Es wird angenommen, dass die Überzeugungen, die Lehrkräfte über die Leistungen individueller Schüler*innen haben, ihr Verhalten gegenüber diesen Lernenden prägen und letztlich die Leistungsentwicklung in Richtung der Erwartungen beeinflussen können. Diese Annahme geht auf das Konzept der selbsterfüllenden Prophezeiung zurück, welches im nächsten Unterkapitel theoriegeleitet beschrieben werden soll. Daran anschließend wird zunächst der empirische Forschungsstand zum Auftreten von Effekten der Leistungserwartungen von Lehrkräften auf die Leistungen der Lernenden zusammengefasst. Abschließend stehen empirische Befunde zum Entstehungsprozess solcher Erwartungseffekte im Zentrum.

2.3.1 Theorie der selbsterfüllenden Prophezeiung inakkurater Lehrkrafterwartungen

Das Phänomen einer selbsterfüllenden Prophezeiung wurde erstmals von Merton (1948, S. 477) beschrieben: „A *false* definition of the situation evoking a new behavior which makes the originally false conception come *true*“. Eine zentrale Voraussetzung für selbsterfüllende Prophezeiungen ist demnach eine basierend auf den aktuell vorliegenden Informationen fehlerhafte Einschätzung der Situation. In interpersonalen Interaktionen kann dies zum Beispiel eine inakkurate Erwartung sein, die

dann zu einem Verhalten beiträgt, das die Annahme über die Situation bzw. die Erwartung letztlich bewahrheitet. Zentral ist hierbei, dass von einer kausalen Beeinflussung ausgegangen wird: Wäre die Einschätzung der Situation ursprünglich anders gewesen, wäre auch das Ergebnis ein anderes gewesen (Biggs, 2009). Selbsterfüllende Prophezeiungen wurden in vielfältigen Kontexten untersucht, zum Beispiel im beruflichen Management oder beim Militär (Dvir, Eden & Banjo, 1995; Eden, 1992; McNatt, 2000).

Die Forschung zu selbsterfüllenden Prophezeiungen im schulischen Kontext geht auf ein weitbekanntes Experiment von Rosenthal und Jacobson (1968) zurück: Lehrkräften wurde über zufällig ausgewählte Schüler*innen der ersten bis sechsten Klassenstufe mitgeteilt, ein durchgeführter Test hätte für diese Kinder diagnostiziert, dass ein überdurchschnittlicher Lernzuwachs im nächsten Jahr von ihnen zu erwarten sei. Obwohl sich die Schüler*innen tatsächlich nicht in ihren Ausgangsleistungen in einem Intelligenztest unterschieden, zeigten diejenigen Kinder, für die ein höherer Lernzuwachs prognostiziert worden war, nach einem Jahr eine bedeutsam höhere Intelligenz als diejenigen Kinder, denen ein solch erhöhter Zuwachs nicht prognostiziert worden war. Dieser Befund wurde als Beleg einer selbsterfüllenden Prophezeiung interpretiert und führte dazu, dass das Phänomen im schulischen Kontext nicht nur unter „Erwartungs- oder Pygmalioneffekt“, sondern auch als „Rosenthal-Effekt“ bekannt wurde. Das Rosenthal-Experiment zog jedoch auch vielfältige Diskussionen über die Verlässlichkeit und Aussagekraft seiner Ergebnisse nach sich. Insbesondere methodische Unzulänglichkeiten wurden kritisiert (für Überblicke zur unmittelbaren Kritik siehe z. B. Jussim & Harber, 2005; Jussim et al., 2009; Lorenz, G., 2018). So seien etwa die Lernzuwächse für das Konstrukt der Intelligenz unplausibel groß ausgefallen (z. B. Snow, 1969; Thorndike, 1968). Folgestudien zeigten, dass diese unplausiblen Intelligenzzuwächse auf statistische Ausreißerwerte zurückgingen, deren Ausschluss die Effekte der Erwartungen so weit reduzierte, dass sie nicht mehr signifikant waren (Snow, 1995). Durch solche und ähnliche Kritik wurden hunderte replizierende experimentelle Studien, Feldstudien und später auch Metaanalysen (z. B. Rosenthal & Rubin, 1978) angestoßen. Die aktuelle empirische Befundlage zum Auftreten von selbsterfüllenden Prophezeiungen im schulischen Kontext wird in Kapitel 2.3.2 genauer dargestellt. Im Kern besteht heutzutage Konsens darüber, dass Erwartungseffekte in der Schule auftreten können (vgl. Jussim et al., 1996; Jussim & Harber, 2005).

Aber wie genau entstehen Erwartungseffekte? Zur Entstehung von Erwartungseffekten existieren sehr viele verschiedene theoretische Modelle (Brophy & Good, 1974; Cooper, 1979; Cooper & Good, 1983; Darley & Fazio, 1980; Jussim, 1986; Rubie, 2003; Trouilloud & Sarrazin, 2003; Wang, Rubie-Davies & Meissel, 2018), die in ihrer Detailliertheit stark variieren, sich über drei zentrale Schritte jedoch einig sind (vgl. Jussim et al., 2009; siehe auch Abb. 3): (1) Zunächst entstehen Erwartungen, die zum Teil inakkurat sind. (2) Diese Erwartungen schlagen sich in differentiell

Handeln nieder. (3) Das differentielle Handeln verändert die Situation so, dass die ehemals inakkuraten Erwartungen bestätigt werden. Dieser Prozess lässt sich in Bezug auf Leistungserwartungen von Lehrkräften und die Leistungsentwicklung von Lernenden folgendermaßen beschreiben: (1) Lehrkräfte bilden Erwartungen über die Leistungspotenziale ihrer Schüler*innen, die den aktuellen Fähigkeiten und lernrelevanten Merkmalen der Lernenden zufolge (zum Teil) inakkurat sind. Die Lehrerin aus unserem Beispiel in Kapitel 2.2.1 könnte z. B. annehmen, Annika könne zwar gut rechnen, tue sich beim Lesenlernen aber sehr schwer und würde daher im kommenden Jahr nur sehr wenig im Lesen dazulernen – obwohl Annika in lesebezogenen Leistungstests eigentlich im Durchschnitt liegt. (2) Die Erwartungen bedingen (unbewusst) das Verhalten der Lehrkraft gegenüber den Schüler*innen. Die Lehrerin von Annika könnte ihr aufgrund ihrer niedrigen Erwartungen also beispielsweise (unbewusst) nur sehr leichte Leseaufgaben geben, die Annika wenig fordern. (3) Auf das von den Erwartungen beeinflusste Verhalten der Lehrkraft reagieren die Lernenden in solch einer Weise, die die ehemals inakkuraten Erwartungen bestätigen. In unserem Beispiel könnte Annika ihre Lesefähigkeiten als Konsequenz der Unterforderung im Schuljahr tatsächlich nur geringfügig verbessern, sodass sich die niedrigen Erwartungen der Lehrerin letztlich bestätigen.

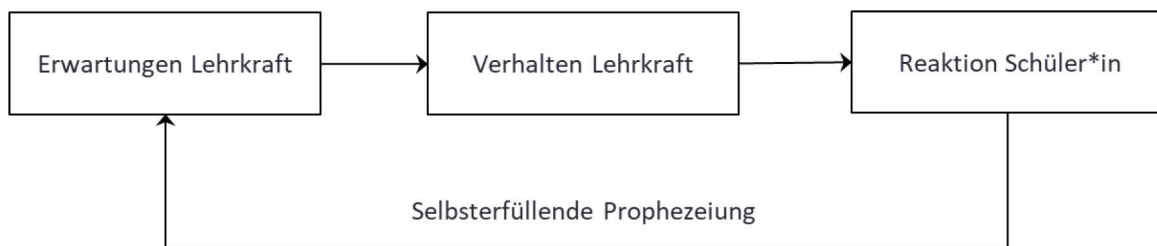


Abb. 3: Drei-Schritte-Modell des Prozesses von Erwartungseffekten (angelehnt an Jussim et al., 2009, S. 361).

Dieses Drei-Schritte-Modell kennzeichnet den Konsens zwischen den verschiedenen vorgeschlagenen Modellen und erlaubt eine einfache Veranschaulichung der angenommenen Prozesse, die zum Entstehen von Erwartungseffekten führen können. Detailliertere Modelle beinhalten genauere Annahmen dazu, wie Erwartungen und Lehrkraftverhalten auf der einen Seite und Lehrkraftverhalten und Änderungen der Leistungen der Schüler*innen auf der anderen Seite miteinander verbunden sein können. Ein solches umfassenderes Modell präsentiert beispielsweise Jussim (1986; siehe Abb. 4), das für ein genaueres Verständnis der angenommen Prozesse kurz beschrieben werden soll. Dieses Modell bündelt vorherige theoretische Überlegungen und empirische Erkenntnisse übersichtlich mit Bezug zu den drei übergeordneten Schritten. Schritt 1 wird dabei um Überlegungen zur Stabilität und Veränderung von Erwartungen ergänzt. Bei Schritt 2 und Schritt 3 integriert das Modell jeweils Annahmen zur Mediation über verschiedene psychologische und

situationale Merkmale. Die einbezogenen Mediatoren orientieren sich dabei eng an der vorherigen Forschung und den bis dahin vorgeschlagenen Modellen. So startet das Modell bei (1) Erwartungen der Lehrkräfte, die von verschiedenen Informationen und Bedingungen beeinflusst werden (u. a. Merkmale der Schüler*innen sowie Stereotype). Diese Erwartungen sind (2) entweder relativ flexibel und werden bei nicht-konformen Verhaltensweisen der Lernenden direkt angepasst oder sie sind relativ stabil und werden auch bei nicht-konformen Verhaltensweisen der Lernenden beibehalten. Stabile inakkurate Erwartungen der Lehrkraft beeinflussen daraufhin vermittelt über (3) einerseits *psychologische Mediatoren auf Seiten der Lehrkraft*, wie etwa die wahrgenommene Kontrolle über Interaktionen mit den Lernenden oder die empfundene Ähnlichkeit und resultierende Sympathie, und über andererseits *situationale Mediatoren*, wie etwa leistungsbezogene Gruppierung oder Klassenstufe, das (4) Verhalten der Lehrkraft auf verschiedenen Dimensionen. Rosenthal (1974) nahm vier Dimensionen relevanten Verhaltens an, in denen sich die Leistungserwartungen der Lehrkräfte niederschlagen könnten. Er vermutete, dass Lehrkräfte Schüler*innen, denen sie höhere Erwartungen entgegenbringen, mehr und anspruchsvollere Arbeitsaufträge erteilen und Materialien zur Verfügung stellen (*Input*), ihnen häufiger und umfangreicher die Möglichkeit geben, ihre Leistungen zu zeigen, zum Beispiel indem sie sie häufiger zu Wort kommen lassen oder nach Fragen länger warten (*Output*), insgesamt einen wärmeren, wertschätzenderen Umgang mit ihnen pflegen (*Klima*) sowie ihnen häufiger positive und detaillierte Rückmeldungen zukommen lassen (*Feedback*). Solche und ähnliche differentielle Lehrkraftverhaltensweisen beeinflussen dem Modell von Jussim (1986) zufolge wiederum (5) *psychologische Mediatoren auf Seiten der Lernenden*. Hier werden beispielsweise die wahrgenommene Kontrolle über Lernerfolge, schulbezogene Überzeugungen, Lernfreude und Motivation sowie Selbstwirksamkeitserwartung und Selbstkonzepte diskutiert. Es wird angenommen, dass sich diese psychologischen Mediatoren wiederum (6) auf der Verhaltensebene der Lernenden niederschlagen und Leistungen sichtbar werden, die die ehemals inakkuraten Erwartungen bestätigen.

Verschiedene Forscher*innen (z. B. Gill & Reynolds, 1999; Kuklinski & Weinstein, 2000) nehmen zudem an, dass es für das Auftreten einer selbsterfüllenden Prophezeiung wichtig sei, dass die Lernenden das differentielle Verhalten der Lehrkraft wahrnehmen und dadurch über die zugrundeliegenden Erwartungen der Lehrkraft informiert werden. In einem neueren Modell ergänzen Wang et al. (2018) daher die von den Schüler*innen wahrgenommenen Lehrkrafterwartungen als Bindeglied zwischen dem tatsächlichen Verhalten der Lehrkraft und den psychologischen Faktoren auf Seiten der Lernenden (z. B. Selbstwirksamkeit oder Selbstkonzept). Ob die Wahrnehmung des Lehrkraftverhaltens durch die Schüler*innen tatsächlich eine notwendige Bedingung für Erwartungseffekte ist und ob sie tatsächlich als Mediator und nicht eher als Moderator fungiert, ist bislang jedoch weitgehend ungeklärt (vgl. Jussim et al., 2009; Wang et al., 2018). In seiner Darstellung

und der Begriffsverwendung geht das Modell von Jussim (1986) bei den benannten psychologischen und situationalen Merkmalen jeweils von *Mediatoren* aus. Gleichzeitig wird allerdings schon bei Jussim (1986) in der genaueren Beschreibung der angenommenen Prozesse und auch in weiterer Forschung zu den Prozessen deutlich, dass diese Merkmale sowohl als *Mediatoren* als auch als *Moderatoren* im Zusammenhang mit Erwartungseffekten stehen könnten. So wird beispielsweise angenommen, dass differentielles Lehrkraftverhalten das akademische Selbstkonzept von Schüler*innen beeinflusst und darüber wiederum die Leistungen bedingt (*Mediation*; z. B. Friedrich et al., 2015). Gleichzeitig werden stärkere Erwartungseffekte für Schüler*innen antizipiert, die ein weniger ausgeprägtes Selbstkonzept haben und noch anfälliger für die Meinungen anderer sind (*Moderation*; z. B. Jussim, 1986; Jussim et al., 1996).

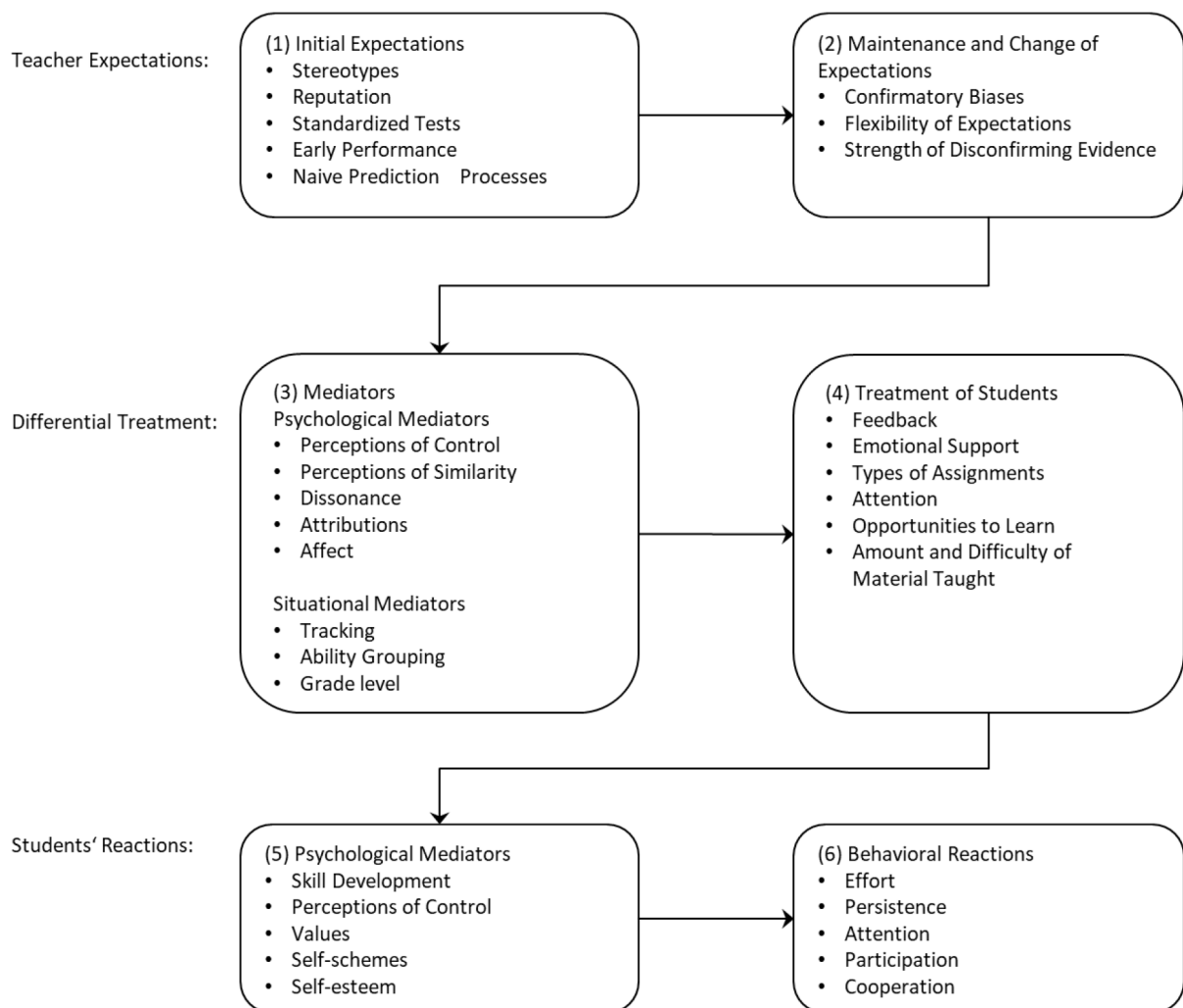


Abb. 4: Detailliertes Modell des Prozesses von Erwartungseffekten nach Jussim (1986, S. 430).

Schon anhand dieser wenigen Beispiele wird deutlich, dass für ein genaueres Verständnis der Entstehung von Erwartungseffekten weitere theoretische Überlegungen und vor allem

prozessbezogene empirische Untersuchungen benötigt werden. Im folgenden Unterkapitel werden zunächst die empirischen Befunde zum Auftreten von Erwartungseffekten im Schulkontext sowie etwaige verstärkende und abschwächende Bedingungen zusammengefasst. Anschließend stehen empirische Befunde zum Vermittlungsprozess über das Lehrkraftverhalten im Blick.

2.3.2 Empirische Befunde zu Erwartungseffekten und ihren Moderatoren

Internationaler Forschungsstand zum Auftreten von Erwartungseffekten

Seit dem klassischen Experiment von Rosenthal und Jacobson (1968; vgl. Kapitel 2.3.1) wurde eine Vielzahl empirischer Studien zum Auftreten von Erwartungseffekten im schulischen Kontext durchgeführt. Der Großteil dieser Studien stammt aus den USA (z. B. Kuklinski & Weinstein, 2001; McKown & Weinstein, 2008; Ready & Chu, 2015; Robinson-Cimpian et al., 2014; Sorhagen, 2013). Aber auch in europäischen Ländern, wie z. B. in den Niederlanden (z. B. de Boer et al., 2010; Jungbluth, 1993, 1994) oder der Schweiz (z. B. Hollenstein et al., 2019), sowie in anderen Regionen der Welt, etwa in Neuseeland (z. B. Peterson et al., 2016) oder China (z. B. Li & Rubie-Davies, 2017; Zhou & Urhahne, 2013), wurden mittlerweile Erwartungseffekte im Schulkontext untersucht. Trotz einer fortdauernden Diskussion um methodische und interpretative Fragen (für nähere Informationen siehe letzter Abschnitt dieses Kapitels) fanden die meisten Studien empirische Belege für Erwartungseffekte, sodass mittlerweile Konsens über das mögliche Auftreten selbsterfüllender Prophezeiungen im schulischen Kontext besteht (vgl. z. B. Alexander & Schofield, 2006; Babad, 2009; Jussim et al., 2009; Wang et al., 2018). Die Effekte fallen allgemein klein aus, wurden aber nicht nur in experimentellen Studien, sondern auch in Feldstudien identifiziert. Korrelativ liegen die Befunde zwischen $r = .10$ und $r = .20$; gemessen in der Effektstärke d betragen sie durchschnittlich $d = .30$ (Jussim et al., 2009). Hattie (2009) kam in seiner Synthese der Ergebnisse von acht Metaanalysen auf eine etwas höhere durchschnittliche Effektstärke von $d = .43$. Das Auftreten von Erwartungseffekte wurde dabei für verschiedene Altersgruppen von Schüler*innen nachgewiesen, wobei neuere Studien primär ältere Lernende am Ende der Primarstufe oder in der Sekundarstufe in den Blick nehmen (z. B. Archambault, Janosz & Chouinard, 2012; de Boer & van der Werf, 2015; Friedrich et al., 2015; Peterson et al., 2016; Zhou & Urhahne, 2013). Weniger Studien betrachten frühere Zeitpunkte in der Bildungslaufbahn. Aber auch in diesen Untersuchungen wurden Belege für das Auftreten von Erwartungseffekten gefunden. So scheinen schon im Elementarbereich die Erwartungen der Betreuungspersonen die mathematischen Fähigkeiten (Schenke, Nguyen, Watts, Sarama & Clements, 2017) und die Literacy-Kompetenzen von Kindergartenkindern (Ready & Chu, 2015) bedeutsam vorherzusagen.

Unklar bleibt dabei bislang, ob positive Effekte hoher Erwartungen („Galatea-Effekte“; Babad, Inbar & Rosenthal, 1982) und negative Effekte niedriger Erwartungen („Golem-Effekte“; Babad et al., 1982) in ähnlicher Stärke auftreten. Madon et al. (1997) fanden in ihrer Studie basierend auf Daten von Sechstklässler*innen stärkere positive Auswirkungen inakkurat hoher Erwartungen als negative Auswirkungen inakkurat niedriger Erwartungen auf die Leistungsentwicklung der Schüler*innen in Mathematik gemessen mit standardisierten Tests. Dieses Befundmuster war besonders ausgeprägt für leistungsstärkere Schüler*innen zu beobachten, bei denen sich keine negativen Erwartungseffekte nachweisen ließen. Im Gegensatz dazu waren in mehreren anderen Studien, die den Einfluss von Erwartungen auf objektive, fächerübergreifende Indikatoren für Schulerfolg untersuchten, die Effektstärken für inakkurat positive und negative Erwartungen vergleichbar (Alvidrez & Weinstein, 1999; de Boer et al., 2010; Rubie-Davies et al., 2014). Deutlich stärkere negative Erwartungseffekte zeigten sich hingegen in der Studie von Alvidrez und Weinstein (1999), wenn der allgemeine Notendurchschnitt (nicht Testleistungen) als abhängige Variable betrachtet wurden, und in der Studie von Babad et al. (1982) zu Auswirkungen von Erwartungen im Sportunterricht für eine von drei untersuchten sportlichen Leistungen. Für die beiden anderen sportlichen Leistungen zeigten sich stärkere positive Erwartungseffekte.

Gleichzeitig sind nicht alle Schüler*innen in allen Klassenkontexten gleichermaßen von positiven und negativen Erwartungseffekten betroffen. Verschiedene individuelle Merkmale der Lernenden, der Lehrkräfte sowie Merkmale des Kontexts scheinen das Auftreten selbsterfüllender Prophezeiungen zu moderieren (vgl. Jussim et al., 1996; Jussim et al., 2009; Wang et al., 2018). Mit Blick auf *individuelle Merkmale der Schüler*innen* weisen manche Studien darauf hin, dass Erwartungseffekte für soziodemografische Gruppen von Schüler*innen, etwa in Abhängigkeit von ihrem sozialen Hintergrund, ethnischen Hintergrund oder Geschlecht, sowie in Abhängigkeit des Leistungsniveaus und Selbstkonzepts der Lernenden unterschiedlich stark ausfallen können. Im Bereich soziodemografischer Merkmale wurden teilweise stärkere Erwartungseffekte auf die Leistungsentwicklung von sozial benachteiligten oder stereotypisierten Gruppen beobachtet, wobei die Befundlage insgesamt eher als gemischt einzustufen ist. So beobachteten Jussim et al. (1996) deutlich stärkere Effekte der Leistungseinschätzungen auf die Mathematikleistungen bei Schüler*innen aus sozial benachteiligten Familien als bei Schüler*innen aus sozial bessergestellten Familien. Auch Hinnant et al. (2009) fanden für Lernende mit niedrigerem sozioökonomischem Status stärkere Erwartungseffekte von der ersten zur dritten Schulklasse in Mathematik. Für Erwartungseffekte von der dritten zur fünften Klasse in Mathematik sowie für beide Altersgruppen im Lesen zeigten sich hingegen keine signifikante Moderation seitens des sozioökonomischen Status der Schüler*innen. De Boer et al. (2010), die Erwartungseffekte über einen Zeitraum von fünf Jahren anhand eines Indikators für den generellen Schulerfolg verfolgt haben, fanden im Kontrast zu den

anderen Studien tendenziell *schwächere* negative Erwartungseffekte bei Schüler*innen mit niedrigem sozioökonomischem Status. Noch deutlicher waren diese Unterschiede wenn die Bildungsaspirationen der Eltern als sozialer Indikator herangezogen wurden: Bei Lernenden, deren Eltern niedrigere Aspirationen zeigten, traten über die fünf Jahre hinweg deutlich schwächere negative Erwartungseffekte auf als bei Lernenden, deren Eltern höhere Aspirationen hatten.

Auch in Abhängigkeit vom ethnischen Hintergrund der Schüler*innen deuten verschiedene Studien auf Unterschiede in der Stärke von Erwartungseffekten hin. In der Studie von Jussim et al. (1996) fiel der Zusammenhang zwischen den Leistungseinschätzungen der Lehrkräfte und den späteren mathematischen Testleistungen sowie Mathematiknoten stärker für afro-amerikanische Schüler*innen aus als für kaukasische Schüler*innen. Dabei traten die Effekte des sozialen und ethnischen Hintergrunds additiv zueinander auf. Auch die Analysen von McKown und Weinstein (2002) unterstützen, dass afro-amerikanische Schüler*innen sowohl in Mathematik als auch im Lesen anfälliger für negative Erwartungseffekte sind. Die Untersuchung von de Boer et al. (2010) fand für die Niederlande hingegen keine bedeutsamen Unterschiede in der Stärke von Erwartungseffekten auf den allgemeinen Schulerfolg in Abhängigkeit vom ethnischen Hintergrund der Schüler*innen. Gemittelt über alle Lernenden hinweg fanden auch Hinnant et al. (2009) weder im Lesen noch in Mathematik Hinweise auf unterschiedlich starke Erwartungseffekte für Schüler*innen in Abhängigkeit von ihrem ethnischen Hintergrund. Jedoch ergaben die Analysen, dass Jungen, die einer ethnischen Minderheit angehören, im Lesen besonders anfällig für Erwartungseffekte von der ersten zur dritten Klassenstufe waren.

Der letztgenannte Befund von Hinnant et al. (2009) weist für den Bereich Lesen und für den Zeitraum von der ersten zur dritten Klassenstufe darauf hin, dass die Stärke von Erwartungseffekten auch vom Geschlecht der Lernenden abhängen kann. Für den Zeitraum von der dritten zur fünften Klassenstufe im Lesen und für beide Zeiträume im Fach Mathematik waren die Mädchen und Jungen der Studie jedoch gleich stark von Auswirkungen der Erwartungen ihrer Lehrkräfte betroffen. Ähnlich inkonsistent fallen die Befunde von Jussim et al. (1996) und McKown und Weinstein (2002) zu geschlechterspezifischen Erwartungseffekten aus. In einzelnen Analysen fanden die Untersuchungen Hinweise auf stärkere Erwartungseffekte für Mädchen in Mathematik, in anderen jedoch nicht. So war die Lehrkräfteeinschätzung der mathematischen Begabung (nicht aber die Einschätzung der mathematischen Leistung oder Anstrengung) bei Mädchen enger mit ihren späteren Mathematiknoten (nicht aber Testleistungen) verknüpft als für Jungen (Jussim et al., 1996). Außerdem waren Mädchen in Mathematik nur dann anfälliger für negative Erwartungseffekte als Jungen, wenn sie zur Gruppe der besonders stark fehleingeschätzten Schüler*innen gehörten, nicht Mädchen generell (McKown & Weinstein, 2002). In ihren Analysen des allgemeinen Schulerfolgs

fanden de Boer et al. (2010) keinen bedeutsamen Unterschied in der Stärke von Erwartungseffekten für Mädchen und Jungen.

Neben soziodemografischen Merkmalen der Schüler*innen wird vermutet, dass das individuelle Leistungsniveau die Anfälligkeit für Erwartungseffekte moderieren kann (vgl. z. B. Madon et al., 1997; Wang et al., 2018). Verschiedene Studien sind dieser Frage nachgegangen und kamen zu unterschiedlichen Ergebnissen. Für den Zeitraum eines Jahres fanden Madon et al. (1997) stärkere Erwartungseffekte auf die mathematischen Leistungen für leistungsschwächere Schüler*innen, wohingegen die Befunde von de Boer et al. (2010) für schwächere Erwartungseffekte auf den allgemeinen Schulerfolg nach fünf Jahren für diese Gruppe von Schüler*innen sprechen. Als möglichen Grund für diese abweichenden Befunde führten de Boer et al. (2010) Unterschiede im untersuchten Zeitraum (ein Jahr vs. fünf Jahre) an. Zusätzliche Analysen für nur ein Jahr bestätigten diese Annahme und zeigten konsistent zu den Befunden von Madon et al. (1997) stärkere Erwartungseffekte für leistungsschwächere Schüler*innen. Auch wenn das akademische Selbstkonzept als abhängige Variable untersucht wird, treten inkonsistente Befunde auf, die möglicherweise mit der Länge des Erhebungszeitraums erklärt werden können. Die Befunde deuten allerdings in die gegensätzliche Richtung. Liu und Wang (2008) fanden in einem Zeitraum von drei Jahren Hinweise auf stärkere Effekte der Lehrkfterwartungen für leistungsschwächere Schüler*innen, wohingegen die Befunde von Pesu, Viljaranta und Aunola (2016) einen engeren Zusammenhang zwischen den von den Lernenden wahrgenommenen Lehrkfterwartungen und ihrem akademischen Selbstkonzept für leistungsstärkere Schüler*innen über den Zeitraum eines Jahres unterstützen.

Auch das akademische Selbstkonzept von Schüler*innen wird als möglicher Moderator der Effekte von Lehrkfterwartungen auf die Leistungsentwicklung von Lernenden diskutiert und empirisch untersucht (vgl. Jussim et al., 1996). Die Befunde von Madon et al. (1997) zeigten stärkere Erwartungseffekte auf die Mathematikleistungen für Schüler*innen mit einem niedrigeren Selbstkonzept als für Schüler*innen mit höherem Selbstkonzept. Diese Unterschiede in den Erwartungseffekten in Abhängigkeit vom Selbstkonzept scheinen allerdings mit der Moderation durch das Leistungsniveau der Lernenden konfundiert zu sein. Wurden beide Interaktionsterme simultan untersucht, war nur die Moderation über die Ausgangsleistung signifikant.

Insgesamt weisen die Ergebnisse darauf hin, dass Erwartungseffekte für verschiedene Gruppen von Schüler*innen unterschiedlich stark ausfallen können, wobei die Befundlage sehr uneinheitlich ist und teilweise gegensätzliche Schlussfolgerungen unterstützt. Wenn auf Lernende mehrere Gruppierungsmerkmale gemeinsam zutreffen, die Erwartungseffekte moderieren, kann dies ihre Anfälligkeit möglicherweise kumuliert verstärken bzw. abschwächen (Jussim et al., 1996; Jussim et al., 2009).

Auch *auf Seiten der Lehrkräfte* werden Merkmale diskutiert, die das Auftreten und die Stärke von Erwartungseffekten beeinflussen können (vgl. Jussim et al., 1996; Wang et al., 2018). Da Erwartungseffekte nur dann auftreten können, wenn inakkurate Erwartungen vorliegen, spielen all jene Merkmale eine Rolle, die beeinflussen, ob Lehrkräfte zu mehr oder weniger akkuraten Urteilen gelangen. Diese Aspekte wurden bereits in Kapitel 2.2.3 beschrieben. Zusätzlich wird angenommen, dass eine stärkere Überzeugung über die Verlässlichkeit der eigenen Urteile sowie eine damit einhergehende Rigidität das Auftreten von selbsterfüllenden Prophezeiungen begünstigt (vgl. Jussim et al., 1996). Werden inakkurate Erwartungen auch dann aufrechterhalten, wenn Lernende wiederholt Leistungen zeigen, die nicht mit den Erwartungen übereinstimmen, erhöht sich die Wahrscheinlichkeit für Erwartungseffekte. Die Lernenden werden dann über einen längeren Zeitraum hinweg immer wieder mit den zu hohen oder zu niedrigen Erwartungen der Lehrkräfte und den darauf basierenden Verhaltensweisen konfrontiert, wodurch sich ihre Leistungen mit höherer Wahrscheinlichkeit den Erwartungen anpassen.

Außerdem wurden Gruppierungen von Lehrkräften anhand des Ausmaßes der Verzerrungen ihrer Urteile und bezüglich ihres Verhaltens gegenüber verschiedenen Gruppen von Schüler*innen vorgenommen und mit der Stärke von Erwartungseffekten in Verbindung gebracht. Babad (1998, 2009) klassifizierte Lehrkräfte danach, wie sehr ihre Beurteilung von Arbeitsergebnissen virtueller Schüler*innen durch experimentell variierte Informationen zu deren sozialen Hintergrund beeinflusst wurden (*high-bias vs. no-bias teachers*). Anschließende Studien zeigten, dass negative Erwartungseffekte in Klassen von *high-bias* Lehrkräften deutlich stärker ausfielen als in Klassen von *no-bias* Lehrkräften, in denen diese quasi nicht vorhanden waren (Babad, 1993; Babad et al., 1982). Dies könnte darauf zurückzuführen sein, dass es *high-bias* Lehrkräften in der Interaktion mit ihren Schüler*innen weniger gut gelingt, negative Emotionen zu verbergen (*nonverbal leakage*), als *no-bias* Lehrkräften (Babad, Bernieri & Rosenthal, 1989a).

Eine ähnliche Klassifikation hat Weinstein (2002) auf der Basis ihrer Forschung zur Schüler*innenwahrnehmung differentiellen Lehrkraftverhaltens entwickelt. In Klassenräumen, in denen die Lernenden stärkere Unterschiede im Verhalten der Lehrkraft gegenüber leistungsstärkeren und leistungsschwächeren Schüler*innen wahrnahmen (*high-PDT; perceived differential treatment*), fanden Kuklinski und Weinstein (2001) teilweise stärkere Erwartungseffekte als in Klassen, in denen die Lernenden den Eindruck hatten, ihre Lehrkraft behandelt alle Lernenden weitgehend gleich (*low-PDT*). So erreichten die Erwartungseffekte in der dritten Klassenstufe in *high-PDT* Klassen eine Effektstärke von $\beta = .35$, während in *low-PDT* Klassen keine Erwartungseffekte auftraten. In der ersten und fünften Klassenstufe unterschied sich die Stärke der Erwartungseffekte hingegen nicht zwischen *high-* und *low-PDT* Klassen.

Auf *kontextueller Ebene* wurden bislang vor allem Unterschiede in Abhängigkeit von der untersuchten Jahrgangsstufe beobachtet. Im ursprünglichen Experiment von Rosenthal und Jacobson (1968) waren die Erwartungseffekte in der ersten und zweiten Schulklasse am stärksten und in der dritten bis sechsten Schulklasse deutlich geringer ausgeprägt. Auch Kuklinski und Weinstein (2001) fanden stärkere Erwartungseffekte in der ersten Klassenstufe als in späteren Klassenstufen des Elementarbereichs. Diese Befunde führten zu der Annahme, dass Erwartungseffekte insbesondere jüngere Kinder betreffen. Wie Jussim und Kolleg*innen (Jussim et al., 1996; Jussim et al., 2009) zusammenfassen, unterstützen die Befunde diese Annahme jedoch nur teilweise. Einer Metaanalyse von Raudenbush (1984) zufolge fielen Erwartungseffekte in der ersten, zweiten und *siebten* Klasse am stärksten aus. Jussim und Kolleg*innen (Jussim et al., 1996; Jussim et al., 2009) schlussfolgern auf dieser Grundlage, dass Erwartungseffekte besonders in neuen, unbekannten Situationen wirksam sein könnten – wie beispielsweise direkt nach der Einschulung, aber auch nach dem Übergang in die weiterführende Schule. Die Ergebnisse einer aktuelleren Studie von de Boer et al. (2010), in der Erwartungseffekte in den ersten fünf Jahren nach dem Übergang in die Sekundarstufe untersucht wurden, stehen im Einklang mit dieser Annahme. Im ersten Jahr waren die Effekte der Lehrkrafterwartungen auf den allgemeinen Schulerfolg der Schüler*innen am stärksten, nahmen im zweiten Jahr ab und blieben danach weitgehend stabil.

Erwartungseffekte in Schulen in Deutschland

Da die vorliegende Dissertation auf Daten aus Deutschland basiert, soll die nationale Befundlage im Folgenden genauer beschrieben werden. Auch in Deutschland werden Auswirkungen von Lehrkrafterwartungen diskutiert (z. B. Ludwig, 2007). Empirische Arbeiten liegen jedoch erst relativ wenige vor. Bislang haben sich fünf Studien mit der Bedeutung individueller Leistungserwartungen oder -einschätzungen für die Leistungsentwicklung von Schüler*innen befasst. Stahl (2007) untersuchte die Lehrkräfteeinschätzung der Lesetestleistung und des Leseinteresses von Drittklässler*innen. Diese individuellen Einschätzungen sagten die Lesetestleistungen und das Leseinteresse der Schüler*innen in der vierten Klassenstufe signifikant positiv vorher, auch wenn die erzielten Leistungen in einem Lesetest und das selbsteingeschätzte Leseinteresse der Kinder zum Zeitpunkt der dritten Klassenstufe kontrolliert wurden. Auch im mathematischen Bereich gingen in einer Studie von Tiedemann (2000) mit höheren Einschätzungen der mathematischen Fähigkeiten durch die Lehrkraft bessere Mathematiknoten von Kindern der dritten und vierten Jahrgangsstufe einher, wenn das Geschlecht des Kindes und die bisherige Mathematiknote berücksichtigt wurde. Ebenfalls mit dem Fach Mathematik befasste sich die Studie von Friedrich et al. (2015). Sechs Monate nach dem Wechsel von der Primar- in die Sekundarstufe gaben die Hauptschullehrkräfte auf zwei

Items an, wie leistungsstark sie jede*n an der Studie teilnehmende*n Schüler*in ihrer fünften Klasse im Fach Mathematik einschätzten. Diese Einschätzung hatte sowohl auf die Mathematiknoten als auch auf die standardisiert erfassten Mathematiktestleistungen am Ende der fünften Klasse einen signifikant positiven Effekt kleiner Effektstärke. Kontrolliert wurden dabei die Mathematiknoten bzw. Mathematiktestleistungen sechs Monate nach Schuljahresbeginn, das Geschlecht und Alter der Jugendlichen sowie ein Maß des schlussfolgernden Denkens. Muntoni und Retelsdorf (2018) analysierten die Bedeutung von geschlechterstereotypen Lehrkrafterwartungen, die wenige Wochen nach dem Wechsel in die fünfte Klasse erhoben wurden, für die Entwicklung der Lesekompetenz von Fünftklässler*innen. Unter Kontrolle der Leseleistungen zu Beginn der fünften Klasse sagten die Leistungserwartungen der Lehrkräfte die Leseleistungen der Schüler*innen in der zweiten Hälfte der sechsten Klasse signifikant vorher. Für das Geschlecht der Lernenden, den sozioökonomischen Status und einen etwaigen Zuwanderungshintergrund der Familien wurde in den Analysen kontrolliert. Lorenz, G. (2018) untersuchte auf der Basis der Daten der KuL-Studie, die auch der vorliegenden Dissertation zugrunde liegen, ob sich in Abhängigkeit vom ethnischen Hintergrund verzerrte Leistungserwartungen von Grundschullehrkräften selbst erfüllen und damit zur Verfestigung ethnischer Bildungsungleichheiten beitragen. Ethnisch verzerrte Leistungserwartungen wurden als Abweichungen der tatsächlich gemessenen Erwartungen von den auf Basis der Ausgangsleistungen und Motivation, dem sozialen Hintergrund und Geschlecht der Schüler*innen vorhergesagten Erwartungen definiert. Diese in Abhängigkeit vom ethnischen Hintergrund der Kinder verzerrten Leistungserwartungen hatten sowohl auf die Entwicklung der Lesekompetenz als auch auf die Entwicklung der Mathematikkompetenz einen signifikant positiven Effekt unter Kontrolle von Ausgangsleistungen und vielfältigen Hintergrundmerkmalen. Da sich in der Stichprobe allerdings keine ethnischen Unterschiede in den Leistungszuwächsen der Kinder über das erste Grundschuljahr nachweisen ließen, die von den Erwartungseffekten hätten erklärt werden können, unterstützen die Befunde den Beitrag von Erwartungseffekten zu ethnischen Bildungsungleichheiten nur bedingt.

Ferner wurden in Studien in Deutschland, die auf Daten einer Längsschnittuntersuchung von Gymnasiast*innen basieren, selbsterfüllende Prophezeiungen der von der Lehrkraft eingeschätzten Eignung zum Erwerb eines Abiturs auf die Wahrscheinlichkeit tatsächlich ein Abitur zu erwerben, ein Studium aufzunehmen und auf das spätere Berufsprestige nachgewiesen (Becker, 2012, 2013; Becker & Birkelbach, 2013; Birkelbach, 2011).

Zusammenfassung und kritische Bewertung

Zusammenfassend lässt sich festhalten, dass das Forschungsinteresse an Erwartungseffekten in den letzten Jahrzehnten sehr groß war und zu einer beachtlichen Anzahl an Studien geführt hat. Die

Befunde dieser Arbeiten unterstützen das Auftreten von Erwartungseffekten im schulischen Kontext mehrheitlich. Vor allem in den USA wurden vielfältige Studien durchgeführt. Für Deutschland liegen bislang nur vereinzelt Ergebnisse dazu vor, ob Erwartungen von Lehrkräften die Leistungsentwicklung von Schüler*innen vorhersagen. Die Stärke von Erwartungseffekten liegt insgesamt im kleinen bis mittleren Bereich, wobei manche Klassen und bestimmte Gruppen von Schüler*innen stärker von Erwartungseffekten betroffen zu sein scheinen als andere.

Die Erforschung von Erwartungseffekten ist allerdings mit Herausforderungen verbunden, die immer wieder für Kritik an den durchgeführten Studien und zu Einschränkungen der Interpretierbarkeit des bisherigen Forschungsstandes führen. Ein zentraler Kritikpunkt bezieht sich auf die Messung von Inakkuratheit in Lehrkrafterwartungen. Es wird kritisiert, dass auftretende Zusammenhänge zwischen Erwartungen und späteren Leistungen dadurch zustande kommen können, dass die Erwartung tatsächliche Kompetenzen der Schüler*innen zutreffend widerspiegeln (z. B. Jussim & Harber, 2005; Timmermans et al., 2015). Eine Herausforderung bei der Analyse von Erwartungseffekten ist also, die Vorhersage der Leistungsentwicklung durch akkurate Erwartungen von Einflüssen inakkurater Erwartungen zu trennen. Nur die letztgenannte Assoziation zwischen inakkuraten Erwartungen und späteren Leistungen kennzeichnet Erwartungseffekte im Sinne einer selbsterfüllenden Prophezeiung (z. B. de Boer et al., 2010). Etwa 40 % der Studien aus den letzten 30 Jahren haben keine Indikatoren für die tatsächlichen Fähigkeiten oder lernrelevanten Merkmale der Schüler*innen in ihre Analysen einbezogen (Wang et al., 2018) und begegnen dadurch dieser Herausforderung nur unzureichend.

Eine weitere Herausforderung bezieht sich auf das Ausmaß des Kontakts zwischen der Lehrkraft und den Schüler*innen, der den Erhebungen vorausging. Häufig werden Erwartungseffekte anhand von Daten zu den Erwartungen der Lehrkräfte und Leistungen der Lernenden untersucht, die erst mehrere Monate oder sogar Jahre (z. B. Friedrich et al., 2015; Hinnant et al., 2009; Peterson et al., 2016; Ready & Chu, 2015) nach Beginn des Lehrkraft-Kind-Kontakts erhoben wurden. In dieser Situation ist es nicht möglich zu unterscheiden, ob die Lehrkrafterwartungen den Leistungen der Schüler*innen entsprechen weil die Lehrkraft die Leistungen präzise eingeschätzt hat oder weil sich die Leistungen durch vorangegangene selbsterfüllende Prophezeiungen an die Erwartungen angepasst haben. Um das Ausmaß von Erwartungseffekten möglichst umfassend untersuchen zu können, sind Studien nötig, die sowohl die Erwartungen der Lehrkräfte als auch die Leistungen der Schüler*innen möglichst bald nach Beginn des Lehrkraft-Kind-Kontakts erfassen. Nur so lässt sich sicherstellen, dass die Erwartungen der Lehrkräfte die Leistungen der Schüler*innen vor deren Erfassung noch nicht beeinflusst hatten und nur deswegen mit ihnen übereinstimmen. Nur in wenigen der bisherigen Studien wurden Erwartungen und Ausgangsleistungen frühzeitig zu Beginn

des Kontakts zwischen Lehrkraft und Schüler*innen erhoben (z. B. Lorenz, G., 2018; Muntoni & Retelsdorf, 2018; Watson et al., 2014).

Zukünftige Studien zu Lehrkrafterwartungen und ihrer Bedeutung für die Leistungsentwicklung von Schüler*innen sollten mit diesen Herausforderungen angemessen umgehen und damit zu einer Absicherung der bisherigen Forschung zum Auftreten von Erwartungseffekten beitragen.

2.3.3 Empirische Befunde zum Prozess von Erwartungseffekten

Während das Auftreten von Erwartungseffekten im Schulkontext mittlerweile gut belegt ist, bleibt weitgehen unklar, wie genau Erwartungen von Lehrkräften die Leistungsentwicklung der Schüler*innen beeinflussen. Eine zentrale Annahme bezieht sich auf das Handeln der Lehrkräfte, welches von den Erwartungen beeinflusst sein soll. Zunächst werden empirische Befunde zum Zusammenhang von Erwartungen und Lehrkraftverhalten zusammengefasst. Daran anschließend wird der Forschungsstand zur Frage, ob das Lehrkraftverhalten die Effekte von Erwartungen auf die Leistungsentwicklung tatsächlich vermittelt, zusammengetragen.

Empirische Befunde zum Einfluss von Erwartungen auf Lehrkraftverhalten

In den 60er, 70er und 80er Jahren fand umfangreiche Forschung zum Zusammenhang von Erwartungen und den von Rosenthal (1974) postulierten vier Dimensionen relevanten Lehrkraftverhaltens – Input, Output, Klima, Feedback (vgl. Kapitel 2.3.1) – statt. Inwiefern die Ergebnisse dieser Forschung die Relevanz der vier Dimensionen unterstützt, untersuchten Harris und Rosenthal (1985) in einer umfassenden Metaanalyse in die 135 Studien einfließen, die bis zum Jahr 1983 durchgeführt wurden. Den Ergebnissen zufolge verhalten sich Lehrkräfte in allen vier Dimensionen unterschiedlich gegenüber Schüler*innen, denen sie höhere Erwartungen entgegenbringen, als gegenüber Schüler*innen, denen sie geringere Fähigkeiten zusprechen. Bringt eine Lehrkraft einem Kind höhere Erwartungen entgegen, ist ihre Interaktion mit dem Kind tendenziell weniger kühl und von geringerer körperlicher Distanz geprägt. Ebenso scheint die Lehrkraft häufiger Blickkontakt mit der*dem Schüler*in zu haben, mehr zu lächeln und insgesamt für ein wärmeres sozioemotionales Klima zu sorgen. Gleichmaßen unterrichten Lehrkräfte Kindern, denen sie höhere Erwartungen entgegenbringen, tendenziell mehr und schwierigere Themen, stellen ihnen häufiger Fragen, interagieren länger und häufiger mit ihnen und die Interaktionen sind stärker auf das Lernen fokussiert. Auch loben die Lehrkräfte diese Kinder tendenziell häufiger, nicken öfter, nehmen Ideen der Schüler*innen auf, ermutigen sie und geben ihnen mehr korrekatives Feedback.

In der Zeit nach Veröffentlichung der Metaanalyse von Harris und Rosenthal (1985) wurden nur sehr wenige Studien zum Zusammenhang zwischen Erwartungen und Lehrkraftverhaltensweisen durchgeführt. Davon basieren manche Studien auf Beobachtungen oder Videografien (z. B. Montague & Rinaldi, 2001; Rubie-Davies, 2007; Wanzek, Roberts & Al Otaiba, 2014), andere beziehen sich auf die Wahrnehmung des Lehrkraftverhaltens durch die Schüler*innen (z. B. Chen, Thompson, Kromrey & Chang, 2011; Matsuoka, 2014; Urhahne, 2015). Auch diese neueren Studien weisen auf Unterschiede im Lehrkraftverhalten gegenüber verschiedenen Gruppen von Schüler*innen hin.

So erhielten Lernende, die ein erhöhtes Risiko für eine Lernschwierigkeit aufwiesen, was häufig mit niedrigeren Lehrkrafterwartungen einhergeht (z. B. Hurwitz et al., 2007; Shifrer, 2013, 2016), negativeres und weniger akademisches Feedback als Lernende ohne ein solches erhöhtes Risiko (Montague & Rinaldi, 2001). Gleichermaßen berichteten Lernende, denen geringere Erwartungen von ihrer Lehrkraft entgegengebracht wurden, sie würden weniger positives und mehr negatives Feedback erhalten (Brattesani, Weinstein & Marshall, 1984; Chen et al., 2011), seltener auf ihr Arbeitsverhalten und das Einhalten von Regeln hingewiesen sowie sich insgesamt weniger wertgeschätzt fühlen und weniger Möglichkeiten erhalten, sich eigenständig am Unterricht zu beteiligen (Brattesani et al., 1984) als ihre Klassenkamerad*innen, denen die Lehrkraft höhere Erwartungen entgegenbrachte. Neuere Beobachtungsstudien stellten allerdings keinen Zusammenhang zwischen den Erwartungen einer Lehrkraft und den Möglichkeiten, die sie ihren Schüler*innen gibt, sich zu äußern oder am Unterricht zu beteiligen, fest (Montague & Rinaldi, 2001; Wanzek et al., 2014). Zu einem ähnlichen Ergebnis gelangte Urhahne (2015) für drei von vier betrachteten Dimensionen von Lehrkraftverhalten aus Sicht der Lernenden selbst. In ihren Leistungen überschätzte und unterschätzte Schüler*innen fühlten sich gleichermaßen von ihren Lehrkräften im Lernen unterstützt, waren der Meinung, dass ihre Lehrkräfte keine Lernenden bevorzugten, und hatten den Eindruck, faire Noten zu erhalten. Allerdings gaben überschätzte Schüler*innen häufiger an, dass ihnen ihre Lehrkraft zugänglich sei und sie mit ihr über alles sprechen könnten als unterschätzte Schüler*innen.

Mit Ungleichbehandlungen von über- und unterschätzten Lernenden auf eher struktureller Ebene befassten sich Ready und Chu (2015). Sie untersuchten den Zusammenhang zwischen inakkuraten Leistungseinschätzungen und der Zuordnung von Kindergartenkindern zu leistungsbezogenen Lesegruppen. Es wurde deutlich, dass die Fachkräfte überschätzte Kinder eher fortgeschrittenen Lesegruppen zuteilten als unterschätzte Kinder.

Babad, Bernieri und Rosenthal (1989b) zeigten, dass unabhängige Rater*innen die Lehrkraftinteraktionen mit Lernenden, denen die Lehrkraft hohe oder niedrige Erwartungen entgegenbrachte, auf der Basis von kurzen Videoclips hinsichtlich des negativen Affekts unterschiedlich einschätzten. In Interaktionen mit Lernenden, denen die Lehrkraft niedrige

Erwartungen entgegenbrachte, wurden demnach insbesondere über non-verbale Kommunikationskanäle mehr negative Emotionen sichtbar als in der Interaktion mit Lernenden, für die die Lehrkraft höhere Erwartungen hatte. Inwieweit sich die Wahrnehmung von differentiellem Lehrkraftverhalten zwischen Lehrkräften und Lernenden unterscheidet, untersuchte auch Babad (1990), der sowohl Lernende als auch Lehrkräfte dazu befragte, ob leistungsstärkere und leistungsschwächere Schüler*innen von der Lehrkraft unterschiedlich behandelt würden. Die Ergebnisse zeigten, dass sich Lernende und Lehrkraft darüber einig waren, dass Leistungsschwächere mehr in ihrem Lernen unterstützt und auf sie weniger Druck ausgeübt würde als auf Leistungsstärkere. Über die Verteilung der emotionalen Unterstützung waren sich Lehrkräfte und Lernende allerdings nicht einig. Während die Lernenden der Meinung waren, dass die Lehrkraft emotional besonders freundlich und unterstützend gegenüber Leistungsstarken sei, gaben die Lehrkräfte an, die Leistungsschwächeren emotional besonders zu unterstützen.

All diese Studien beziehen sich auf die Zusammenhänge individueller Erwartungen von Lehrkräften und dem auf individuelle Schüler*innen bezogenen Lehrkraftverhalten. Eine weitere Gruppe aktueller Studien, die an dieser Stelle nur kurz angesprochen werden soll, bezieht sich auf die Unterrichtspraxis von Lehrkräften, die allen Schüler*innen ihrer Klasse hohe Erwartungen entgegenbringen (*high-expectancy teachers*), im Vergleich zu der Unterrichtspraxis von Lehrkräften mit niedrigeren Erwartungen an die Lernenden ihrer Klasse (*low-expectancy teachers*). Rubie-Davies und Kolleg*innen haben in einer Reihe von Studien Merkmale identifiziert, in denen sich die Unterrichtspraxis dieser beiden Gruppen von Lehrkräften unterscheidet. Lehrkräfte, die allen Schüler*innen relativ hohe Erwartungen entgegenbrachten, gestalteten insgesamt eine günstigere Lernumgebung, gaben mehr Feedback, stellten den Lernenden häufiger Fragen, die komplexere Denkprozesse anstoßen, und zeigten ein besseres Klassenmanagement als *low-expectancy* Lehrkräfte (z. B. Rubie-Davies, 2007).

Ein weiteres Forschungsfeld bezieht sich auf die Frage, ob sich Schüler*innen die Erwartungen ihrer Lehrkräfte durch deren Verhalten erschließen können. In verschiedenen Studien konnten Babad und Kolleg*innen (z. B. Babad, Bernieri & Rosenthal, 1991; Babad & Taylor, 1992; für einen Überblick siehe Babad, 2009) eindrucksvoll zeigen, dass bereits junge Kinder anhand von 10-sekündigen Videoclips sehr erfolgreich identifizierten, ob eine Lehrkraft ihrem Gegenüber hohe oder niedrige Leistungserwartungen entgegenbringt (Babad et al., 1991). Die Studien zeigten weiterhin, dass diese Entschlüsselung auch dann noch erfolgreich war, wenn lediglich non-verbale Informationen zur Verfügung standen (Babad & Taylor, 1992).

Zusammenfassend wird deutlich, dass sich Lehrkräfte basierend auf ihren Erwartungen recht unterschiedlich gegenüber ihren Schüler*innen verhalten. Dabei spielen verschiedenste Verhaltensweisen eine Rolle, die nicht nur verbale, sondern auch subtilere Formen der

Kommunikation und Interaktion betreffen. Gleichzeitig gelingt es Lernenden recht gut, aus dem Verhalten einer Lehrkraft ihre zugrundeliegenden Erwartungen abzulesen. Damit haben bisherige Studien verschiedene Verhaltensweisen der Lehrkräfte identifiziert, die möglicherweise als Bindeglied zwischen den Erwartungen der Lehrkraft und Änderungen in Leistungsauscomes der Schüler*innen fungieren. Zu beachten bleibt hierbei allerdings, dass fast alle berichteten Studien den absoluten Zusammenhang zwischen Erwartungen und Lehrkraftverhaltensweisen (unabhängig vom tatsächlichen Leistungsniveau der Schüler*innen) betrachtet haben (Ausnahmen sind Ready & Chu, 2015; Rubie-Davies, 2007; Urhahne, 2015). Für Erwartungseffekte im Sinne einer selbsterfüllenden Prophezeiung müssten sich allerdings *inakkurate* Erwartungen in Lehrkraftverhalten niederschlagen, das wiederum die Entwicklung der Schüler*innen beeinflusst. Die hier berichteten Lehrkraftverhaltensweisen können daher als Pool möglicher Verhaltensmediatoren verstanden werden, deren tatsächlicher Beitrag zur Entstehung von Erwartungseffekten noch zu klären ist. Hierbei geht es zum einen darum zu klären, ob die differenziellen Verhaltensweisen lediglich auf den tatsächlichen Ausgangsleistungen der Schüler*innen basieren, oder ob sie auch mit inakkuraten Erwartungen assoziiert sind. Zum anderen gilt es zu untersuchen, ob bzw. zu welchem Anteil die Verhaltensweisen die Erwartungseffekte tatsächlich medieren. Der Forschungsstand zur Mediation wird im Folgenden beschrieben.

Empirische Befunde zur Mediation der Erwartungseffekte durch Lehrkraftverhalten

Die wichtigste Studie, die frühe Forschung zur Mediation von Erwartungseffekten durch das Lehrkraftverhalten bündelt, ist die bereits oben beschriebene Metaanalyse von Harris und Rosenthal (1985). In dieser Metaanalyse wurden nicht nur Befunde zum Zusammenhang von Erwartungen und verschiedenen Lehrkraftverhaltensweisen untersucht, sondern auch Befunde zu deren Assoziation mit Outcomes auf Seiten der Schüler*innen. Für alle vier Verhaltensdimensionen (Input, Output, Klima und Feedback) erbrachte die Metaanalyse Hinweise auf deren Relevanz. Die Dimensionen Input und Klima zeigten über alle Studien und zugeordneten Verhaltensweisen hinweg die stärksten Zusammenhänge sowohl mit den Erwartungen der Lehrkräfte (Input: $r = .26$, Klima: $r = .20$) als auch mit den Outcomes auf Seiten der Schüler*innen (Input: $r = .33$, Klima: $r = .37$). Die Korrelationen für die Dimension Output (Erwartungen Lehrkräfte: $r = .19$, Outcomes Schüler*innen: $r = .20$) und Feedback (Erwartungen Lehrkräfte: $r = .13$, Outcomes Schüler*innen: $r = .07$) fielen geringer aus. In späteren Arbeiten überführte Rosenthal (1989) das 4-Faktoren-Modell daher in ein reduziertes 2-Faktoren-Modell, das die besondere Relevanz der beiden Faktoren Klima und Input – neu benannt als *affect* und *effort* – hervorheben sollte. Das 4-Faktoren-Modell blieb dennoch in der Forschung zu Erwartungseffekten populär.

Betrachtet man die Ergebnisse der Metaanalyse (Harris & Rosenthal, 1985) im Detail, fällt eine teilweise große Variation der Zusammenhänge auf. Insbesondere die einzelnen Indikatoren, die in der Metaanalyse zur *Feedback* Dimension zusammengefasst wurden, zeigten wenig einheitliche, teilweise gegensätzliche Zusammenhangsmuster mit den Lehrkfterwartungen und Leistungen der Schüler*innen. So variierten die identifizierten Zusammenhänge zwischen Erwartungen und Feedback zwischen $r = -.05$ und $r = .36$. Die über alle Indikatoren hinweg kombinierte Effektstärke betrug $r = .13$. Für den Zusammenhang zwischen Feedback und Schüler*innenoutcomes war diese Variation mit kombinierten Effektstärken von $r = -.23$ bis $r = .12$ sogar noch deutlicher. Der mittlere Effekt für den Zusammenhang von Feedback und Leistungen lag bei $r = .07$. Dies wiederum resultierte in einer Schätzung für den indirekten Effekt von ungefähr null.

Trotz der wichtigen Einblicke, die die Metaanalyse zum Zusammenhang von Erwartungen und Lehrkraftverhalten gewährt, stößt ihre Aussagekraft mit Blick auf die Mediation von Erwartungseffekten über das Lehrkraftverhalten durch zwei Limitationen an ihre Grenzen. Zum einen stellten Harris und Rosenthal (1985) nicht sicher, dass die in die Metaanalyse einbezogenen Studien Indikatoren für die tatsächlichen Leistungen der Schüler*innen berücksichtigten. Entsprechend lässt sich nicht entscheiden, ob die gefundenen Assoziationen auch für inakkurat zu hohe oder zu niedrige Erwartungen zutreffen. Zum anderen haben nur wenige der einbezogenen Studien sowohl den Link zwischen Erwartungen und Verhalten als auch den Link zwischen Verhalten und Leistungen untersucht. Demzufolge konnte die Metaanalyse der Frage nach der Mediation nicht direkt nachgehen (vgl. Chow, 1987).

Zwei neuere Studien haben die Mediation der Effekte von Erwartungen über Lehrkraftverhalten auf Outcomes der Schüler*innen direkt untersucht. Urhahne (2015) ging der Frage nach, ob das von den Lernenden wahrgenommene Lehrkraftverhalten in vier Dimensionen die Erwartungseffekte auf die Motivation und Emotionen der Schüler*innen mediiert. Überschätzte Schüler*innen fühlten sich von ihrer Lehrkraft besser emotional unterstützt als unterschätzte Schüler*innen (vgl. auch weiter oben in diesem Kapitel). Diese schülerseitig wahrgenommenen Unterschiede im Lehrkraftverhalten vermittelten die Erwartungseffekte auf die Motivation und Emotionen der Lernenden teilweise. Ready und Chu (2015) untersuchten, ob die Leistungsgruppierung durch Lehrkräfte den Zusammenhang zwischen ihren Erwartungen und der Leseleistung von Kindergartenkindern vermittelt. Überschätzte Kinder wurden dabei von den Lehrkräften eher leistungsstärkeren Gruppen zugeteilt als unterschätzte Kinder (vgl. auch weiter oben in diesem Kapitel). Diese differentielle Zuteilung zu Lerngruppen auf der Grundlage inakkurater Erwartungen stellte sich allerdings nur als schwacher Mediator der Erwartungseffekte heraus.

Weitere aktuelle Studien, in denen eine potentielle Mediation von Erwartungseffekten im Schulkontext durch das Verhalten der Lehrkräfte direkt untersucht wurde, sind mir nicht bekannt.

Auch Wang et al. (2018) kommen auf der Grundlage eines umfangreichen Reviews von Studien der letzten 30 Jahre zu dem Schluss, dass nur wenige Studien zum Zusammenhang von Erwartungen und Lehrkraftverhalten durchgeführt wurden und keine dieser wenigen Studien zusätzlich auch den Zusammenhang des Lehrkraftverhaltens mit der Leistungsentwicklung untersucht hat.

Zusammenfassend lässt sich also festhalten, dass der Prozess von Erwartungseffekten im Schulkontext noch weitgehend ungeklärt ist. Zwar liegen aus bisherigen Studien wichtige Befunde zu möglicherweise relevanten Verhaltensweisen der Lehrkräfte vor. Ob diese Verhaltensweisen die Erwartungseffekte allerdings tatsächlich vermitteln, bleibt bislang unklar. Zudem sind die vorliegenden Befunde zur Rolle von Feedback uneinheitlich.

III Ziele und Fragestellungen der Arbeit

Übergeordnetes Ziel der vorliegenden Arbeit ist, zu einem genaueren Verständnis darüber beizutragen, auf welchen Merkmalen von Schüler*innen Erwartungen und Einschätzungen von Lehrkräften über Leistungen von Lernenden basieren und über welche Prozesse sie das schulische Lernen von Schüler*innen beeinflussen. Dabei wird angenommen, dass Leistungserwartungen und -einschätzungen von Lehrkräften als ein Aspekt professioneller Handlungskompetenz eine wichtige Grundlage für die Gestaltung individuell angepasster Lernangebote und einer lernförderlichen Lehrkraft-Kind-Interaktion darstellen. Ein genaueres Verständnis über die Grundlage von Leistungserwartungen und -einschätzungen und ihre Bedeutung für Lernprozesse ist daher sowohl für eine optimale Förderung individueller Schüler*innen wichtig als auch zur Realisierung gleicher Bildungschancen für Kinder verschiedener soziodemografischer Gruppen bedeutsam. Für erfolgreiches pädagogisches Handeln werden Erwartungen und Einschätzungen als wichtig erachtet, die Leistungen und lernrelevante Merkmale von Schüler*innen akkurat widerspiegeln (vgl. Kapitel 2.2.3). Mit Blick auf existierende ethnische, soziale und geschlechtsbezogene Bildungsungleichheiten ist die Frage zentral, ob Erwartungen und Einschätzungen von Lehrkräften nicht nur von den tatsächlichen Fähigkeiten der Lernenden, sondern auch von ihrem ethnischen Hintergrund, sozialen Hintergrund oder Geschlecht beeinflusst werden. Dass dies teilweise der Fall sein kann, belegen internationale Studien, wobei die Befundlage insgesamt inkonsistent ist. Zudem liegen für Deutschland bislang nur sehr wenige Arbeiten zu Verzerrungen in Leistungserwartungen und -einschätzungen vor (vgl. Kapitel 2.2.4). Daher untersucht *Teilstudie 1* (Lorenz, G., Gentrup, Kristen, Stanat & Kogan, 2016), ob Leistungserwartungen von Lehrkräften zu Beginn der ersten Grundschulklasse in Abhängigkeit des ethnischen Hintergrunds, sozialen Hintergrunds oder Geschlechts der Lernenden verzerrt sind. Abbildung 5 fasst die in Teilstudie 1 zu untersuchenden Zusammenhänge zusammen. Zunächst wird derjenige Anteil der Leistungserwartungen bestimmt, der mit den Leistungen, den allgemeinen kognitiven Fähigkeiten und der Motivation der Schüler*innen kongruiert (akkurate Leistungserwartungen). Bestehen unter Kontrolle dieser Merkmale Unterschiede in den Leistungserwartungen in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund oder Geschlecht der Lernenden, wäre dies ein Hinweis auf Verzerrungen. Diese Unterschiede werden in Abbildung 5 durch den Zusammenhang zwischen den Hintergrundmerkmalen der Schüler*innen und dem inakkuraten Anteil der Erwartungen dargestellt. Um auch Informationen über mögliche Unterschiede für verschiedene ethnische Herkunftsgruppen zu erlangen, werden die größten Zuwanderungsgruppen in den Berechnungen unterschieden. Auf der Grundlage bisheriger Befunde zu Leistungserwartungen und dem vergleichsweise schlechten Abschneiden von Kindern mit türkischem

Zuwanderungshintergrund in Schulleistungsstudien (z. B. PISA: Gebhardt, Rauch, Mang, Sälzer & Stanat, 2013; Stanat, Rauch & Segeritz, 2010) werden insbesondere für Kinder dieser Zuwanderungsgruppe niedrigere Leistungserwartungen prognostiziert. Ebenfalls wird angenommen, dass Lehrkräfte vergleichsweise niedrige Erwartungen für Kinder aus sozial benachteiligten Familien sowie für Mädchen in Mathematik und für Jungen im Lesen haben.

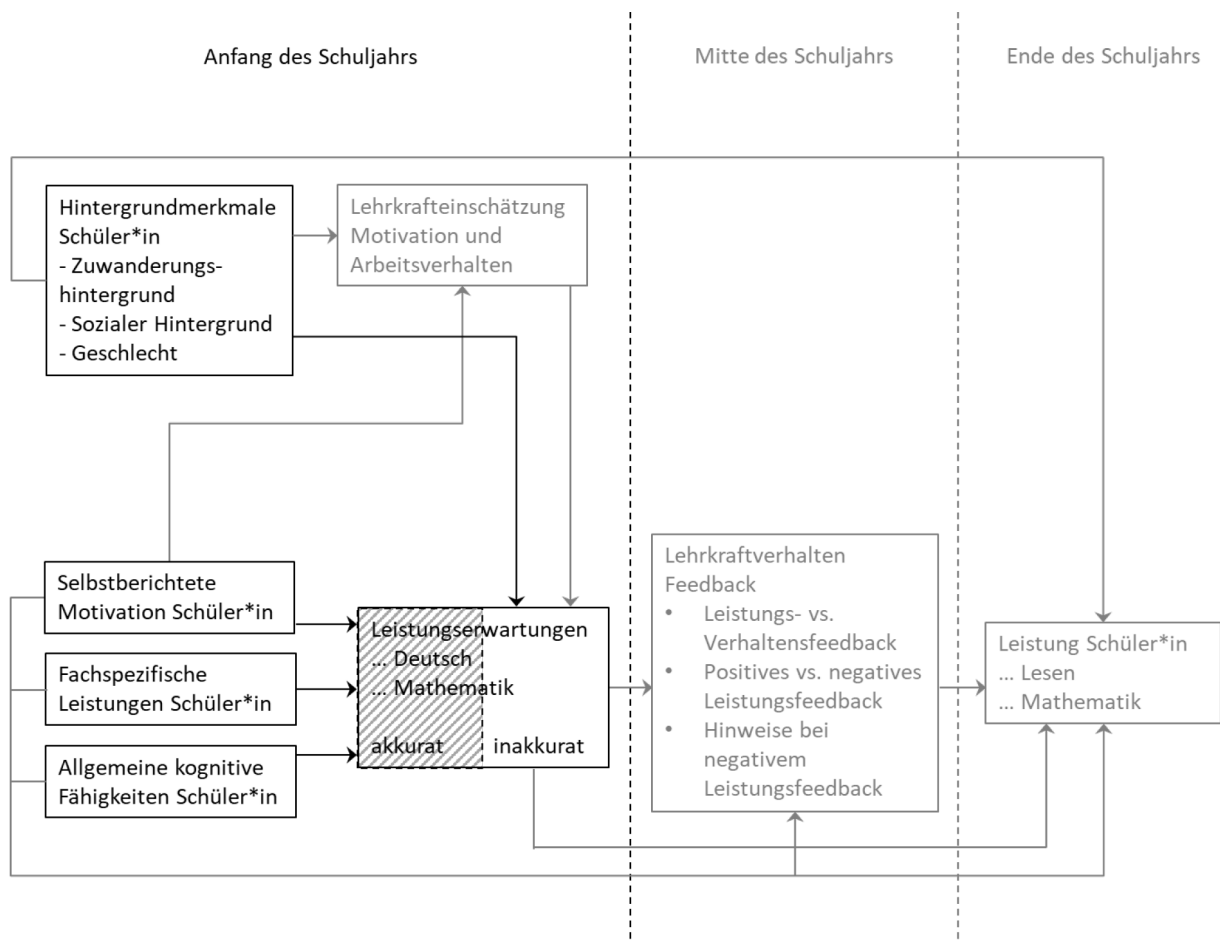


Abb. 5: Überblick über Teilstudie 1.

Sozialpsychologischer Literatur zufolge können Verzerrungen in Leistungserwartungen auf kategorienbasierte Informationsverarbeitungsprozessen und damit auf den Einfluss inakkuratere Stereotypen zurückzuführen sein (vgl. Kapitel 2.2.2). Solche stereotypen Annahmen können sich auf verschiedene Merkmale der Lernenden beziehen. Neben Annahmen über die schulischen Leistungen der Gruppen kommen u. a. auch Überzeugungen zu ihrem Lernverhalten und zu ihrer Motivation in Betracht (vgl. Kapitel 2.2.5). Mit dieser möglichen Ursache von Verzerrungen in Leistungserwartungen befasst sich *Teilstudie 2* (Gentrup, Rjosk, Stanat & Lorenz, 2018) und untersucht die Bedeutung von Lehrkrachteinschätzungen der Motivation und des Arbeitsverhaltens von Lernenden. Eine grafische Darstellung der Teilstudie 2 ist in Abbildung 6 zu finden. Zunächst wird der Frage nachgegangen, ob

ethnische, soziale und geschlechtsbezogene Verzerrungen in Leistungserwartungen auf die Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens von Lernenden zurückgeführt werden können. Diese Beziehung ist in Abbildung 6 als indirekter Pfad von den Hintergrundmerkmalen der Schüler*innen über die Lehrkrachteinschätzungen der Motivation und des Arbeitsverhaltens zum inakkuraten Anteil der Leistungserwartungen der Lehrkräfte dargestellt. Wenn die Verzerrungen vollständig auf Unterschiede in der Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens zurückgeführt werden können, würde sich der direkte Pfad von den Hintergrundmerkmalen der Schüler*innen zu den Leistungserwartungen dem Wert Null annähern. Zusätzlich wird untersucht, zu welchem Anteil die Einschätzungen der Lehrkräfte Selbsteinschätzungen der Kinder widerspiegeln, um Hinweise darüber zu erlangen, ob es sich eher um akkurate Einschätzungen oder stereotypenbasierte Annahmen über die Motivation und das Arbeitsverhalten handelt. Es wird erwartet, dass sich die Lehrkrachteinschätzungen der Motivation und des Arbeitsverhaltens der Schüler*innen über deren Selbsteinschätzung hinaus in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund und Geschlecht der Lernenden unterscheiden. Konkret wird angenommen, dass die Lehrkräfte Kindern aus sozial bessergestellten Familien bzw. Mädchen eine höhere Motivation und ein günstigeres Arbeitsverhalten zusprechen als Kindern aus sozial benachteiligten Familien bzw. Jungen. Hinsichtlich des ethnischen Hintergrunds lässt sich auf der Grundlage des theoretischen und empirischen Forschungsstands keine gerichtete Hypothese formulieren. Weiterhin wird angenommen, dass sich Verzerrungen in Leistungserwartungen von Lehrkräften zum Teil auf Unterschiede in der Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens zurückführen lassen.

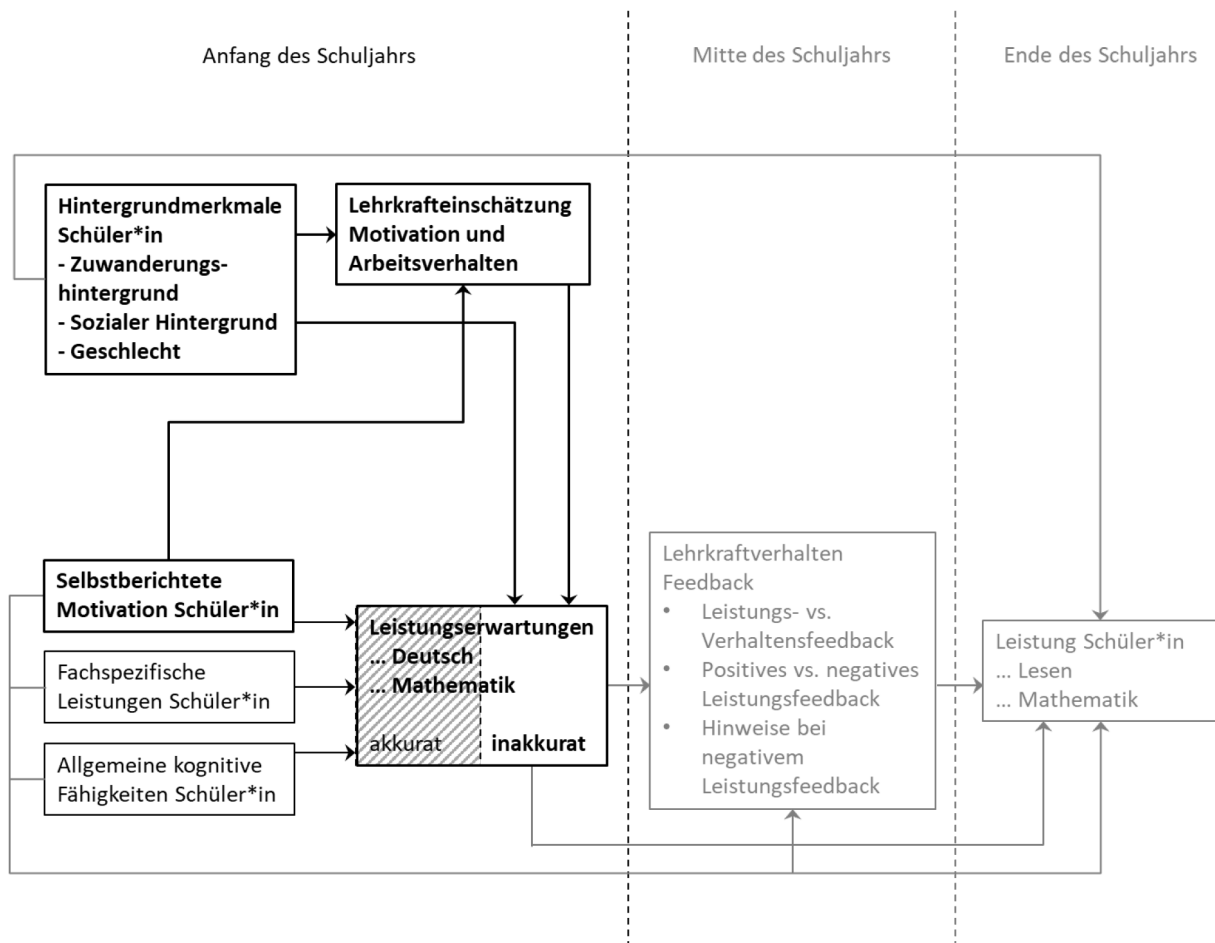


Abb. 6: Überblick über Teilstudie 2.

Die ersten beiden Teilstudien der vorliegenden Dissertation rücken die Urteilsbildung in den Blick und geben Aufschluss über die Genauigkeit und mögliche Verzerrungen von Leistungserwartungen sowie Hinweise zu deren Ursprung. Die Relevanz dieser Fragestellungen ist mit der Annahme verbunden, dass Leistungserwartungen von Lehrkräften die weitere Lernentwicklung von Schüler*innen beeinflussen können. Das Auftreten solcher Erwartungseffekte ist mittlerweile gut belegt, wenngleich in Deutschland noch recht wenige Studien zu dieser Frage existieren (vgl. Kapitel 2.3.2). Weitgehend ungeklärt ist allerdings, über welche Prozesse genau die Leistungserwartungen von Lehrkräften die Leistungsentwicklung der Lernenden bedingen (vgl. Kapitel 2.3.3). *Teilstudie 3* (Gentrup, Lorenz, Kristen & Kogan, 2020) greift dieses Desiderat auf und untersucht den Prozess von Erwartungseffekten (siehe Abb. 7). Dazu wird zunächst die Replizierbarkeit internationaler Befunde für den deutschen Schulkontext überprüft und analysiert, ob Erwartungen von Lehrkräften die Leistungsentwicklung von Erstklässler*innen in Grundschulen in Deutschland im Sinne einer selbsterfüllenden Prophezeiung beeinflussen. In Abbildung 7 ist diese Annahme durch den direkten Pfad von inakkuraten Leistungserwartungen zu den Leistungen der Schüler*innen dargestellt. Die Hintergrundmerkmale der Schüler*innen und ihre Leistungen, allgemeinen kognitiven Fähigkeiten

und motivationalen Merkmale zu Schuljahresbeginn dienen als Kontrollvariablen. Es wird erwartet, dass sich auch für den deutschen Schulkontext Erwartungseffekte nachweisen lassen und entsprechend inakkurat niedrige Erwartungen nach Berücksichtigung der Kontrollvariablen mit geringeren Leistungszuwächsen und inakkurat hohe Erwartungen mit höheren Leistungszuwächsen assoziiert sind. Anschließend wird der angenommene 3-Schritte-Prozess von Erwartungen über Lehrkraftverhalten – spezifisch der Feedbackpraxis – hin zu Auswirkungen auf die Leistungsentwicklung der Schüler*innen in den Blick genommen. Ausgehend von den uneinheitlichen Ergebnissen zum Feedback von Lehrkräften, wird diese Dimension des Lehrkrafthandelns fokussiert. Es wird untersucht, ob das Feedback der Lehrkräfte Effekte der Erwartungen auf die Leistungen der Erstklässler*innen vermittelt. Dazu wird angenommen, dass inakkurat hohe Erwartungen mit mehr und positiverem Leistungsfeedback assoziiert sind, das im Fall eines Fehlers eher Hinweise zur Unterstützung des weiteren Lernens enthält, als inakkurat niedrigere Erwartungen. Dieses differentielle Feedback sollte zudem die Effekte inakkurater Leistungserwartungen auf die Leistungsentwicklung der Schüler*innen zum Teil medieren.

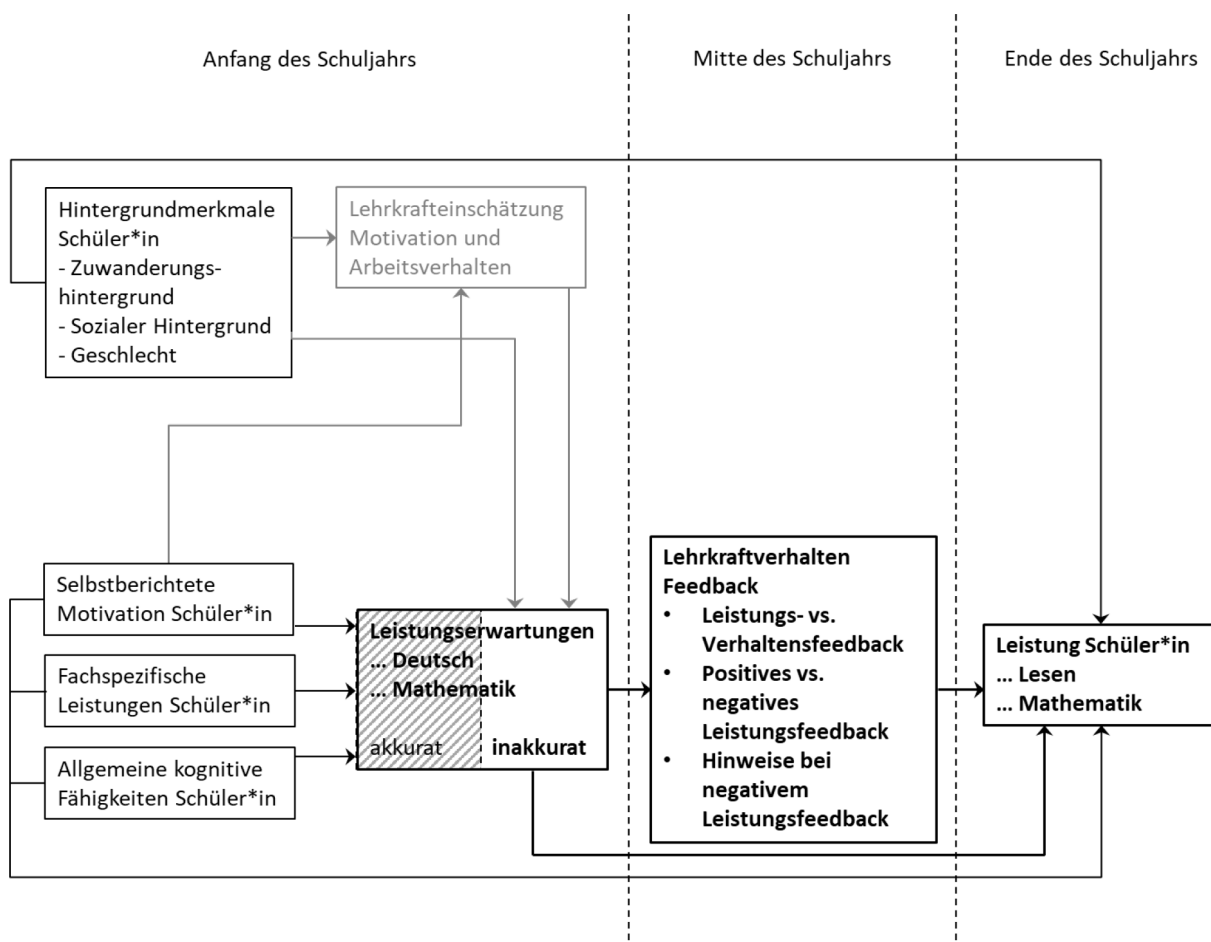


Abb. 7: Überblick über Teilstudie 3.

Die Datengrundlage aller drei Teilstudien stammt aus dem Forschungsprojekt „Kompetenzerwerb und Lernvoraussetzungen“ (KuL)⁵, das im Schuljahr 2013/2014 an 39 Grundschulen in Nordrhein-Westfalen durchgeführt wurde (Kristen et al., 2018a, 2018b). In drei Erhebungswellen begleitete das Projektteam die insgesamt 67 ersten Grundschulklassen über den Zeitraum von einem Jahr. Die erste Erhebungswelle begann wenige Wochen nach der Einschulung, die zweite Welle fand in der Mitte des ersten Schuljahrs und die dritte Erhebungswelle am Ende des ersten Schuljahrs statt. Die Erhebungen umfassten Kompetenzmessungen und Befragungen der Schüler*innen, Lehrkraftfragebögen und telefonische Elterninterviews. Ein Teil der Schulklassen nahm darüber hinaus an der optionalen Videostudie in der Mitte des Schuljahrs teil. Das Design der KuL-Studie ist durch viele Vorteile für die Untersuchung von Erwartungseffekten gekennzeichnet. Ein zentraler Aspekt bezieht sich auf den frühen Beginn der Studie. Da bereits wenige Wochen nach der Einschulung die Ausgangsfähigkeiten der Erstklässler*innen sowie die Erwartungen der Lehrkräfte erfasst wurden, war zum Erhebungszeitpunkt der vorangegangene Lehrkraft-Kind-Kontakt relativ gering. Dies reduziert das Risiko, dass die Erwartungen der Lehrkräfte die Leistungen der Kinder bereits vor der Erhebung in Richtung der Erwartungen beeinflusst haben und dadurch die Akkuratheit der Erwartungen überschätzt werden könnte. Ein weiterer Vorteil liegt in der inhaltlichen Breite der Erhebungen. Sowohl die erfassten Erwartungen und Einschätzungen der Lehrkräfte als auch die gemessenen Kompetenzen und Vorläuferfähigkeiten der Schüler*innen beziehen sich auf die Bereiche Lesen und Mathematik und decken somit zwei für den weiteren Bildungserfolg zentrale Kompetenzbereiche ab. Gleichzeitig erlauben die vielfältigen weiteren lernrelevanten Merkmale und Hintergrundinformationen, die für die Erstklässler*innen erhoben wurden, eine umfangreiche Berücksichtigung von Drittvariablen in den jeweiligen Analysen. Zu guter Letzt stellt die zusätzliche Videostudie eine wichtige Stärke der Studie dar. Diese ermöglicht es, die individuellen Interaktionen zwischen Lehrkraft und Schüler*innen zu analysieren und damit den zugrundeliegenden Prozessen von Erwartungseffekten nachzugehen.

⁵ Das Projekt wurde vom Bundesministerium für Bildung und Forschung gefördert (Förderkennzeichen: 01JC1117A, B, C) und unter der Leitung von Prof. Dr. Cornelia Kristen (Otto-Friedrich-Universität Bamberg), Prof. Dr. Irena Kogan (Universität Mannheim) und Prof. Dr. Petra Stanat (Humboldt-Universität zu Berlin) durchgeführt.

IV Teilstudien der Arbeit

4.1 Teilstudie 1

Lorenz, G., Gentrup, S., Kristen, C., Stanat, P. & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen [Stereotypes among teachers? A study of systematic bias in teacher expectations]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(1), 89–111. Doi: 10.1007/s11577-015-0352-3

Abstract: Die vorliegende Studie geht der Frage nach, ob Lehrererwartungen unabhängig von den Kompetenzen der Schüler nach dem Geschlecht, nach der sozialen oder nach der ethnischen Herkunft variieren. Systematische Verzerrungen liegen dann vor, wenn Lehrkräfte die zukünftigen Leistungen in Abhängigkeit askriptiver Merkmale unter- oder überschätzen. Derartige Muster könnten durch unzutreffende Stereotype zustande kommen. Vor diesem Hintergrund werden die Leistungserwartungen von 69 Grundschullehrkräften untersucht. Die Ergebnisse belegen, dass die zu Schuljahresbeginn geäußerten Leistungserwartungen im Fach Deutsch bei gleichen Ergebnissen in objektiven Leistungstests für Kinder mit einem türkischen Zuwanderungshintergrund, für Kinder aus sozial schwächeren Familien sowie für Jungen negativ verzerrt sind. Bei der Betrachtung des Unterrichtsfachs Mathematik finden sich positive Verzerrungen für Kinder mit einem osteuropäischen Zuwanderungshintergrund und für Kinder aus sozial besser gestellten Familien. Geschlechtsspezifische Fehleinschätzungen im Fachbereich Mathematik sind nicht erkennbar.

4.2 Teilstudie 2

Gentrup, S., Rjosk, C., Stanat, P. & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen [Teachers' perceptions of students' motivation and learning behaviour and their role in biased teacher achievement expectations]. *Zeitschrift für Erziehungswissenschaft*, 21, 867–891. Doi: 10.1007/s11618-018-0806-2

Abstract: Verschiedene Studien haben gezeigt, dass Leistungserwartungen von Lehrkräften teilweise in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund sowie vom Geschlecht der Schülerinnen und Schüler stereotyp verzerrt sein können. Im vorliegenden Beitrag wird die Annahme untersucht, dass diese systematischen Verzerrungen in Leistungserwartungen im sprachlichen und

mathematischen Bereich teilweise auf die Lehrkrifteinschätzung der Motivation und des Arbeitsverhaltens der Schülerinnen und Schüler zurückgeführt werden können. Die Ergebnisse, die auf einer Stichprobe von $N = 901$ Kindern aus $N = 66$ ersten Grundschulklassen mit 69 Klassen- und Fachlehrkräften basieren, unterstützen dies für soziale und geschlechtsbezogene Verzerrungen. Verzerrungen in Abhängigkeit vom ethnischen Hintergrund der Kinder scheinen hingegen unabhängig von der Lehrkrifteinschätzung der Motivation und des Arbeitsverhaltens der Schülerinnen und Schüler aufzutreten. Rückschlüsse auf stereotype Annahmen und deren Bedeutung für unterrichtliches Handeln sowie für Bildungsgleichheit werden diskutiert.

4.3 Teilstudie 3

Gentrup, S., Lorenz, G., Kristen, C. & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, 66. Doi: 10.1016/j.learninstruc.2019.101296

Abstract: This study investigated the effects of teacher expectations on student learning, relying on longitudinal data from 64 classrooms and 1026 first-grade students in Germany. Further, based on a subsample encompassing 19 classrooms with 354 students, we explored the mediating role of three characteristics of teacher feedback rated in video-recorded school lessons. The results showed that teacher expectations were inaccurate to some extent; that is, they did not entirely agree with students' current achievement, general cognitive abilities and motivations. In addition, this inaccuracy in teacher expectations significantly predicted students' end-of-year achievement, even after considering prior achievement, general cognitive abilities, motivation, and student background characteristics. Specifically, inaccurately high teacher expectations were associated with greater achievement in reading and mathematics, whereas inaccurately low teacher expectations were associated with lower achievement in reading only. Furthermore, teacher feedback varied significantly with inaccurate teacher expectations but did not substantially mediate teacher expectancy effects.

V Gesamtdiskussion

5.1 Zusammenfassung der Befunde und Einordnung in den Forschungsstand

Die vorliegende Dissertation ging Fragen zur Bildung von Erwartungen und Einschätzungen von Lehrkräften und ihrer Bedeutung für die Leistungsentwicklung von Schüler*innen nach (siehe Abb. 8 für ein Gesamtmodell der Dissertation). Im Folgenden sollen die empirischen Erkenntnisse der Arbeit zu den Teilfragestellungen zusammengefasst, in den bisherigen Forschungsstand eingeordnet und mit Bezug auf angrenzende Forschungsfelder diskutiert werden. Das Kapitel ist anhand der beiden übergeordneten Themenbereiche der Dissertation gegliedert: (1) Urteilsbildung und (2) Erwartungseffekte.

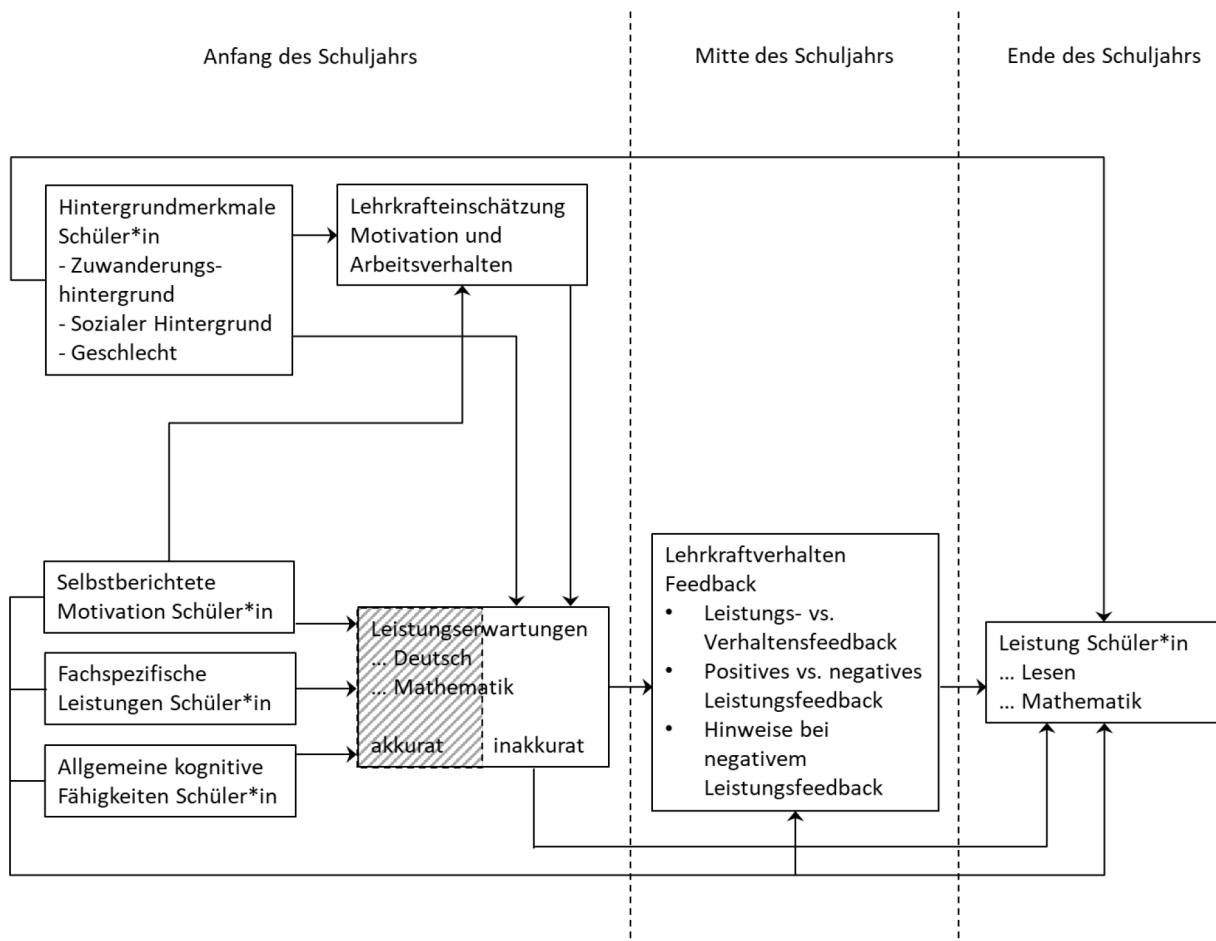


Abb. 8: Arbeitsmodell der vorliegenden Dissertation.

5.1.1 Urteilsbildung – (In)Akkuratheit von Erwartungen und Einschätzungen sowie Einflüsse von Stereotypen

Es wurde untersucht, in welchem Ausmaß Erwartungen von Lehrkräften für die Leistungen von Erstklässler*innen mit lernrelevanten Merkmalen der Kinder übereinstimmen und ob sich die Erwartungen in Abhängigkeit von soziodemografischen Merkmalen der Kinder unterscheiden. Die Ergebnisse der Teilstudien 1 und 3 bestätigen die Annahme, dass Leistungserwartungen von Lehrkräften substantiell mit lernrelevanten Merkmalen der Schüler*innen assoziiert sind, aber ein bedeutsamer Anteil der Erwartungen dennoch unerklärt bleibt. Diese Abweichungen von beobachteten lernrelevanten Merkmalen traten nicht zufällig, sondern systematisch für verschiedene Gruppen von Schüler*innen auf (Teilstudie 1). Im sprachlichen Bereich wurden die Leistungen von Kindern mit türkischem Zuwanderungshintergrund, aus Familien mit niedrigerem sozioökonomischem Status und von Jungen systematisch unterschätzt. In Mathematik zeigten sich positive Verzerrungen in den Leistungserwartungen für Kinder mit osteuropäischem Zuwanderungshintergrund und für Kinder aus sozial bessergestellten Familien. Geschlechterspezifische Unterschiede in den Leistungserwartungen, die über gemessene Unterschiede in den lernrelevanten Merkmalen der Kinder hinausgingen, traten in Mathematik nicht auf. Weiterführende Analysen in Teilstudie 2 verdeutlichten weiterhin, dass Lehrkräfte auch in der Motivation und im Arbeitsverhalten von Schüler*innen Unterschiede in Abhängigkeit von deren Hintergrundmerkmalen antizipierten. Diese Unterschiede kongruierten nur zu einem sehr kleinen Anteil mit Selbsteinschätzungen der Erstklässler*innen und könnten demnach für Verzerrungen in Leistungserwartungen mitverantwortlich sein. Die Befunde von Teilstudie 2 bestätigten diese Annahme für Verzerrungen in Abhängigkeit vom sozialen Hintergrund und Geschlecht. Verzerrungen in Abhängigkeit vom ethnischen Hintergrund scheinen unabhängig von der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens von Schüler*innen aufzutreten. Im Folgenden werden die Befunde der Teilstudien zur (In)Akkuratheit von Leistungserwartungen sowie zu Unterschieden und Verzerrungen in Abhängigkeit der drei untersuchten Hintergrundmerkmale nacheinander genauer beschrieben, in den bestehenden Forschungsstand eingeordnet und diskutiert.

Ausmaß der (In)Akkuratheit von Leistungserwartungen und Zusammenhänge mit Einschätzungen der Motivation und des Arbeitsverhaltens

Die Ergebnisse der Teilstudien 1 und 3 verdeutlichen, dass 35 bis 40 % der Variation in den Leistungserwartungen von Lehrkräften mit den fachspezifischen Leistungen der Schüler*innen zu Schuljahresbeginn, ihren allgemeinen kognitiven Fähigkeiten und ihrer selbsteingeschätzten

schulischen Motivation kongruieren. Basierend auf den theoretischen Überlegungen zu validen Informationsquellen für Leistungserwartungen (vgl. Kapitel 2.1.2) wird dieser Teil in der vorliegenden Arbeit als akkurat beurteilt (zum möglichen Einfluss von Messfehlern oder unbeobachteten Schüler*innenmerkmalen siehe Kapitel 5.2). Umgekehrt ließen sich 60 bis 65 % der Unterschiede in den Leistungserwartungen nicht durch die genannten lernrelevanten Merkmale der Schüler*innen erklären. Diese Ergebnisse zum Ausmaß der (In)Akkuratheit stimmen mit Ergebnissen von Metaanalysen zur Kongruenz von Leistungsurteilen und anhand von standardisierten Tests erfassten Leistungen der Lernenden überein (Hoge & Coladarci, 1989; Südkamp et al., 2012). Verglichen mit den Befunden von de Boer et al. (2010; geteilte Varianz: 67 %), die zur Vorhersage der Übergangsempfehlungen auch motivationale Merkmale der Lernenden einbezogen, fällt der in der vorliegenden Dissertation als akkurat identifizierte Anteil allerdings geringer aus. Als Grund könnte man Unterschiede in der Kontaktdauer zwischen den Lehrkräften und Kindern annehmen, da die vorliegenden Analysen auf Daten zu neu eingeschulten Kindern basieren, wohingegen sich die Befunde von de Boer et al. (2010) auf den Übergang von der Primar- in die Sekundarstufe beziehen. Demnach konnten die Lehrkräfte in der Studie von de Boer et al. zum Zeitpunkt der Einschätzung wahrscheinlich auf umfangreichere Informationen über und Erfahrungen mit den Lernenden zurückgreifen. Gegen diese Annahme sprechen allerdings zum einen Befunde bisheriger Studien aus dem Bereich der diagnostischen Kompetenz, denen zufolge kein substantieller Zusammenhang zwischen der Kontaktdauer und der Urteilsakkuratheit besteht (z. B. Oerke et al., 2016; Wild & Rost, 1995; vgl. auch Kapitel 2.2.3). Außerdem variierte das Ausmaß der (In)Akkuratheit in Teilstudie 1 nicht mit dem Datum des Rückversands des Lehrkräftefragebogens, welches zwischen der dritten Schulwoche und Anfang März variierte, wobei über 75 % der Fragebögen bis zur siebten Schulwoche zurückgeschickt wurden (Median = zweite Schulwoche). Eine größere Bedeutung für die Unterschiede im Ausmaß der (In)Akkuratheit dürfte hingegen der antizipierten Relevanz des Urteils zukommen. Messen Personen einem Urteil eine höhere Relevanz zu, fällt das Urteil tendenziell auch akkurater aus (vgl. Kapitel 2.2.2). Die Urteilsrelevanz dürfte aus Sicht der Lehrkräfte für Übergangsempfehlungen um einiges höher sein als für Leistungserwartungen, die in einem Fragebogen im Rahmen einer Studie freiwillig angegeben werden.

Die erklärten Varianzanteile in Leistungserwartungen variierten leicht zwischen Teilstudie 1 (Deutsch: 37 %, Mathematik: 36 %) und Teilstudie 3 (Deutsch: 34 %, Mathematik: 37 %). Dies liegt zum einen darin begründet, dass sich die Operationalisierung der Leistungserwartungen zwischen den Teilstudien leicht unterschied. Während in Teilstudie 1 ein einzelnes Item für die Messung der Erwartungen herangezogen wurde, wurden in Teilstudie 3 die Leistungserwartungen anhand mehrerer Items abgebildet. Außerdem unterschieden sich die Analysestrategien zwischen den Studien dahingehend, dass in Teilstudie 1 Mehrebenenregressionsmodelle und in Teilstudie 3

Regressionsmodelle mit für die Mehrebenenstruktur korrigierten Standardfehlern berechnet wurden. Entsprechend bezieht sich der erklärte Varianzanteil in Teilstudie 1 auf die Variation auf der Ebene der Schüler*innen (Level 1), während es sich in Teilstudie 3 um erklärte Anteile in der Gesamtvariation handelte. Die trotz dieser methodischen Unterschiede nur leichten Abweichungen in den erklärten Varianzanteilen weisen auf eine Robustheit der Befunde hin.

Des Weiteren deuten die Ergebnisse der Teilstudie 2 auf einen engen Zusammenhang zwischen Leistungserwartungen und Lehrkrachteinschätzungen der Motivation und des Arbeitsverhaltens von Schüler*innen hin, was mit Ergebnissen bisheriger Forschung zu Leistungseinschätzungen übereinstimmt (z. B. Anders, McElvany, et al., 2010; Trautwein & Baeriswyl, 2007; vgl. Kapitel 2.2.5). Demzufolge hat eine Lehrkraft für Lernende, denen sie eine hohe Motivation und ein günstiges Arbeitsverhaltens zuspricht, auch höhere Erwartungen an ihre Leistungsentwicklung. In Erweiterung des Forschungsstands belegen die Ergebnisse von Teilstudie 2 zudem, dass die Assoziation unabhängig von den gemessenen Leistungen, allgemeinen kognitiven Fähigkeiten, motivationalen Merkmalen und Hintergrundmerkmalen der Kinder besteht. Dieses Ergebnis steht im Einklang mit Befunden von Timmermans et al. (2016), die den Zusammenhang von Lehrkrachteinschätzungen des Arbeitsverhaltens mit Übergangsempfehlungen untersucht haben, und mit den Ergebnissen von Robinson-Cimpian et al. (2014) zu Leistungseinschätzungen im Fach Mathematik.

Ethnischer Hintergrund – Unterschiede und Verzerrungen in Erwartungen und Einschätzungen sowie Rückschlüsse auf Stereotype

Die Befunde aus Teilstudie 1 und 2 verdeutlichen, dass sich sowohl Erwartungen von Lehrkräften für die Leistungen von Schüler*innen als auch Einschätzungen ihrer Motivation und ihres Arbeitsverhaltens für verschiedene ethnische Gruppen von Schüler*innen unterscheiden und dies zum Teil über die beobachtbaren Unterschiede zwischen den Gruppen hinausgeht. Hervorzuheben ist hierbei, dass in den Analysen der soziale Hintergrund und das Geschlecht kontrolliert waren, sodass es sich spezifisch um Verzerrungen in Abhängigkeit vom ethnischen Hintergrund handelt. Bisherige internationale und nationale Studien hatten zwar ebenfalls Unterschiede in den Leistungserwartungen von Lehrkräften für Kinder mit und Kinder ohne Zuwanderungshintergrund identifiziert, diese Unterschiede stimmten jedoch mehrheitlich mit beobachtbaren Leistungsdifferenzen überein (z. B. Ready & Wright, 2011; Rjosk et al., 2011; van den Bergh et al., 2010). Dennoch stehen die vorliegenden Befunde nicht im Widerspruch zum bisherigen Forschungsstand. In den vorliegenden Analysen wurden, anders als in bisherigen Studien, verschiedene ethnische Herkunftsgruppen getrennt voneinander untersucht. Dabei zeigte sich, dass

die Unterschiede in den Erwartungen und Einschätzungen spezifisch für verschiedene Herkunftsgruppen auftreten und zum Teil sogar in unterschiedliche Richtungen weisen. Solche gegensätzlichen Effekte für verschiedene Herkunftsgruppen könnten sich in bisherigen Studien ausgemittelt haben.

Für die Gruppe der Kinder mit *türkischem Zuwanderungshintergrund* fielen die Leistungserwartungen in Deutsch und Mathematik insgesamt niedriger aus als für Kinder ohne Zuwanderungshintergrund, wobei die Erwartungen im sprachlichen Bereich beobachtete Leistungsunterschiede überschätzten. Entsprechend zeigten sich negativ verzerrte Erwartungen für Kinder mit türkischem Zuwanderungshintergrund im Fach Deutsch, nicht aber im Fach Mathematik. Diese Ergebnisse reihen sich in die uneinheitliche Befundlage neuerer experimenteller Studien mit Lehrkräften in Deutschland ein, die teilweise negative Verzerrungen (Glock & Krolak-Schwerdt, 2013; Holder & Kessels, 2017; Tobisch & Dresel, 2017), teilweise keine Verzerrungen (Glock & Krolak-Schwerdt, 2014) für Kinder mit türkischem Zuwanderungshintergrund in verschiedenen fachlichen Domänen identifiziert hatten (vgl. auch Kapitel 2.2.4). Zusammengenommen deuten die Studienergebnisse mit einer leichten Tendenz darauf hin, dass diese Schüler*innengruppe eher im sprachlichen als im mathematischen Bereich unterschätzt wird. Ein systematisches Muster, demzufolge negative Verzerrungen ausschließlich im sprachlichen Bereich auftreten, lässt sich aber in der Forschung im Allgemeinen nicht eindeutig identifizieren.

Schulleistungstudien wiesen in der Vergangenheit darauf hin, dass die Gruppe der Schüler*innen mit türkischem Zuwanderungshintergrund häufig schulisch weniger erfolgreich ist als ihre Klassenkamerad*innen ohne Zuwanderungshintergrund (z. B. für 15-Jährige in den PISA-Studien: Gebhardt et al., 2013; Segeritz, Walter & Stanat, 2010; Stanat et al., 2010). So zeigten etwa im IQB-Bildungstrend 2016 Viertklässler*innen, von denen ein Elternteil oder beide Elternteile in der Türkei geboren sind, geringere Kompetenzen im Lesen und in Mathematik als Viertklässler*innen ohne Zuwanderungshintergrund (Rjosk et al., 2017). Die Leistungsnachteile von Kindern, deren Eltern beide in der Türkei geboren sind, fielen dabei im Lesen tendenziell stärker aus als in Mathematik (Viertklässler*innen im IQB-Bildungstrend 2016: Rjosk et al., 2017; 15-Jährige in den PISA-Studien: Gebhardt et al., 2013; Stanat et al., 2010). Solche und ähnliche Befunde werden öffentlichkeitswirksam berichtet. Zudem dürften sich diese Leistungsverteilungen zu einem gewissen Grad auch in einzelnen Schulklassen und Kohorten von Schüler*innen widerspiegeln und damit die Erfahrungen von Lehrkräften prägen. Beides könnte sich in stereotypen Annahmen von Lehrkräften über verschiedene Schüler*innengruppen niederschlagen. Dass dies tatsächlich der Fall sein könnte, legen die Ergebnisse der vorliegenden Arbeit nahe, da die Richtung der von den Lehrkräften angenommenen Leistungsunterschiede mit den in Schulleistungstudien berichteten Disparitäten übereinstimmen. Die akkuraten Erwartungen an die Mathematikleistungen der Schüler*innen mit

türkischem Zuwanderungshintergrund können dabei sowohl auf die Verarbeitung individueller Informationen im Rahmen einer eigenschaftsbasierten Informationsverarbeitung als auch auf die Anwendung eines Stereotyps, das die Gruppenunterschiede in den Mathematikleistungen zutreffend widerspiegelt, zurückgehen (siehe zu akkuraten und inakkuraten Stereotypen auch Kapitel 2.2.2). Die Leistungserwartungen im Fach Deutsch an Kinder mit türkischem Zuwanderungshintergrund scheinen hingegen von einem inakkuraten Stereotyp, das die Leistungsunterschiede zwischen den Gruppen überschätzt, beeinflusst zu sein (worauf die negativen Verzerrungen im Fach Deutsch hindeuten). Gleichzeitig scheint kein spezifisches Stereotyp zur Anstrengungsbereitschaft oder Motivation von Schüler*innen mit türkischem Zuwanderungshintergrund zu bestehen, da Lehrkräfte Kinder mit türkischem Zuwanderungshintergrund und Kinder ohne Zuwanderungshintergrund gleich motiviert und anstrengungsbereit einschätzten, was auch deren Selbsteinschätzungen entspricht. Die Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens erklärten entsprechend auch nicht die negativ verzerrten Erwartungen im sprachlichen Bereich für diese Herkunftsgruppe. Stattdessen könnten die geringeren Erwartungen der Lehrkräfte im sprachlichen Bereich beispielsweise auf der Annahme geringerer Deutschkenntnisse basieren. In einer Interviewstudie von Kratzmann und Pohlmann-Rother (2012) thematisierten beispielsweise alle befragten Erzieherinnen geringe Deutschkenntnisse von Kindern mit türkischem Zuwanderungshintergrund, die ihrer Einschätzung nach zu Herausforderungen im Kindergartenalltag führen.

Während sich die Ergebnisse zu Kindern mit türkischem Zuwanderungshintergrund relativ gut anhand theoretischer Überlegungen und bisheriger Forschung erklären lassen, überraschen die Befunde zu Kindern mit *osteuropäischem Zuwanderungshintergrund*⁶. Diese Gruppe von Schüler*innen wurde von den Lehrkräften vergleichsweise positiv eingeschätzt. In allen drei untersuchten Bereichen fanden sich höhere Erwartungen und Einschätzungen für Kinder mit osteuropäischem Hintergrund als für Kinder ohne Zuwanderungshintergrund; die Lehrkräfte hatten also höhere Erwartungen im sprachlichen und mathematischen Bereich an ihre Leistungen und nahmen sie als motivierter und anstrengungsbereiter wahr als ihre Mitschüler*innen ohne Zuwanderungshintergrund. Während sich die höheren Erwartungen im sprachlichen Bereich mit beobachteten Leistungsvorteilen dieser Gruppe deckten, deuten die Befunde in Mathematik auf positive Verzerrungen hin. Diese positiv verzerrten Leistungserwartungen in Mathematik konnten nicht mit der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens erklärt werden, obwohl

⁶ In dieser Gruppe von Schüler*innen wurden in den Teilstudien 1 und 2 Kinder zusammengefasst, von denen mindestens ein Großeltern- oder Elternteil in einem der folgenden Staaten geboren ist: alle europäischen und nicht-europäischen Staaten der ehemaligen Sowjetunion sowie Polen, Tschechien, Bulgarien, Rumänien, Ungarn, Slowakei und Slowenien. Den größten Anteil dieser Gruppe machten Schüler*innen aus, deren Familien aus Polen (etwa 60 %) oder aus Staaten der ehemaligen Sowjetunion (etwa 30 %) zugewandert sind.

Kinder mit osteuropäischem Hintergrund von den Lehrkräften auch als motivierter und anstrengungsbereiter eingeschätzt wurden.

Die vergleichsweise positiven Leistungserwartungen für Kinder mit osteuropäischem Zuwanderungshintergrund lassen sich mit Ergebnissen von Schulleistungstudien nur bedingt erklären. Zwar erzielen sie im Vergleich zu Kindern mit türkischem oder anderem Zuwanderungshintergrund tendenziell bessere Ergebnisse; einen Vorteil im Vergleich zu Schüler*innen ohne Zuwanderungshintergrund belegen die Studien allerdings nicht (z. B. Viertklässler*innen im IQB-Bildungstrend: Rjosk et al., 2017; 15-Jährige bei PISA: Segeritz et al., 2010). Stattdessen fielen im Jahr 2016 die Leistungen im Lesen und in Mathematik von Viertklässler*innen, deren Eltern in der ehemaligen Sowjetunion geboren sind, nach Kontrolle ihres sozialen Hintergrunds vergleichbar zu den Leistungen ihrer Klassenkamerad*innen ohne Zuwanderungshintergrund aus. Ein ähnliches Bild zeigte sich für die Viertklässler*innen, deren Eltern in Polen geboren sind, wobei für Kinder mit beiden Elternteilen aus Polen auch nach Kontrolle des sozialen Hintergrunds noch ein leichter Nachteil im Lesen zu verzeichnen war (Rjosk et al., 2017). Einen zahlenmäßigen, aber nicht signifikanten Vorteil in Mathematik für 15-jährige Schüler*innen der zweiten Generation mit Eltern aus der ehemaligen Sowjetunion fanden Segeritz et al. (2010) auf Basis der PISA-E-2003-Daten nach Kontrolle des sozialen Hintergrunds. Insgesamt liefern diese Befunde einen Grund für den vergleichsweise günstigen Status der Schüler*innen mit osteuropäischem Hintergrund verglichen mit Kindern mit einem anderen Zuwanderungshintergrund. Die positive Verzerrung in den Leistungserwartungen im Fach Mathematik im Vergleich zu Kindern ohne Zuwanderungshintergrund lässt sich anhand dieser Ergebnisse jedoch nicht begründen. Möglicherweise spielen unterschiedliche Vergleichsgruppen eine Rolle. Lehrkräfte könnten Kinder mit osteuropäischem Zuwanderungshintergrund als zugehörig zur Gruppe der Kinder mit Zuwanderungshintergrund empfinden, innerhalb dieser Gruppe aber als vergleichsweise leistungsstark wahrnehmen. Dies könnte dazu führen, dass durchschnittliche oder gute Leistungen in dieser Schüler*innengruppe besonders positiv wahrgenommen werden, was letztlich in einer besonders starken Überschätzung resultieren würde. Warum die positiven Verzerrungen allerdings nur im mathematischen und nicht im sprachlichen Bereich auftreten, lässt sich damit nicht begründen. Stattdessen könnten die inakkurat hohen Erwartungen der Lehrkräfte im mathematischen Bereich auf antizipierte Werthaltungen der Eltern basieren. Lehrkräfte könnten bei den Eltern der Kinder mit osteuropäischem Hintergrund eine besondere Wertschätzung der mathematisch-naturwissenschaftlichen Fächer wahrnehmen (z. B. Walter & Taskinen, 2008), was sich wiederum in inakkurat hohen Erwartungen in Mathematik für diese Kinder niederschlagen könnte.

Sozialer Hintergrund – Unterschiede und Verzerrungen in Erwartungen und Einschätzungen sowie Rückschlüsse auf Stereotype

Für alle drei untersuchten Erwartungen und Einschätzungen der Lehrkräfte zeigte sich eine bedeutsame Assoziation mit dem sozialen Hintergrund der Schüler*innen. Bereits zu Beginn der ersten Grundschulklasse nahmen die Lehrkräfte an, dass Kinder aus sozial bessergestellten Familien eine höhere Motivation und ein günstigeres Arbeitsverhalten zeigen und sich in Deutsch und Mathematik besser entwickeln werden als Kinder aus sozial benachteiligten Familien. Dabei überschätzten die Lehrkräfte beobachtete Unterschiede zwischen den sozialen Gruppen von Schüler*innen bei allen drei Merkmalen. Nur ein sehr kleiner Teil der Einschätzungen der Motivation und des Arbeitsverhaltens konnte auf Selbsteinschätzungen der Kinder zurückgeführt werden. Für die Leistungen in Deutsch und Mathematik nahmen die Lehrkräfte etwa doppelt so große soziale Unterschiede an als sie auf Basis der gemessenen Ausgangsleistungen, allgemeinen kognitiven Fähigkeiten und motivationalen Merkmalen der Lernenden zu erwarten wären. Sowohl die Richtung der antizipierten Unterschiede als auch die fächerübergreifende Vergleichbarkeit der Befunde stimmt mit Ergebnissen großangelegter Schulleistungsstudien überein. Zum einen belegten diese Studien wiederholt, dass die erzielten Leistungen im Lesen (z. B. Viertklässler*innen bei IGLU 2016: Hußmann, Stubbe & Kasper, 2017; 15-Jährige bei PISA 2009: Ehmke & Jude, 2010) und in Mathematik (z. B. Viertklässler*innen bei TIMSS 2015: Stubbe, Schwippert & Wendt, 2016; 15-Jährige bei PISA 2012: Müller & Ehmke, 2013) mit dem sozialen Hintergrund der Schüler*innen zusammenhängen, wobei Schüler*innen aus sozial bessergestellten Familien bessere Leistungen erzielten. Zum anderen fallen die sozialen Leistungsdisparitäten im Lesen und in Mathematik ähnlich groß aus, was etwa die durch den HISEI erklärten Varianzanteile in den Leistungen von 14 bzw. 13 % verdeutlichen (Viertklässler*innen im IQB-Bildungstrend 2016: Haag, Kocaj, Jansen & Kuhl, 2017).

Der Befund zu sozialen Verzerrungen in Leistungserwartungen steht im Einklang mit der internationalen Forschung, die mehrheitlich auch nach Kontrolle beobachteter Leistungsunterschiede geringere Leistungserwartungen und -einschätzungen für Kinder aus sozial benachteiligten Familien identifizierte (z. B. Ready & Wright, 2011; Timmermans et al., 2016; vgl. auch Kapitel 2.2.4). Für Deutschland stimmen die vorliegenden Befunde mit den Ergebnissen der drei Feld- und Vignettenstudien überein, die ebenfalls soziale Verzerrungen identifizierten (Becker, 2013; Stahl, 2007; Tobisch & Dresel, 2017), wohingegen sie im Widerspruch zu den Ergebnissen der anderen drei Feld- und Vignettenstudien stehen, die keine sozialen Verzerrungen gefunden haben (Glock & Krolak-Schwerdt, 2014; Hachfeld et al., 2010; Karing et al., 2011).

Die Richtung und Fachunabhängigkeit der sozialen Verzerrungen in Leistungserwartungen stimmt mit dem Stereotyp überein, Kinder aus sozial schlechtergestellten Familien seien allgemein

schwächer in der Schule. Dass sich dieses Stereotyp nicht (oder zumindest nicht ausschließlich) auf ein geringeres Vorwissen oder eine fehlende Begabung bezieht, verdeutlicht der Befund, dass die sozialen Verzerrungen in Leistungserwartungen etwa zur Hälfte auf die Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens zurückgeführt werden konnten. Die Lehrkräfte scheinen Kinder aus sozial benachteiligten Familien als weniger motiviert und anstrengungsbereit wahrzunehmen und für sie unter anderem deshalb auch eine weniger günstige Leistungsentwicklung zu prognostizieren als für Kinder aus sozial bessergestellten Familien. Da Lehrkräfte für Kinder aus sozial bessergestellten Familien aber auch dann noch höhere Erwartungen hatten, wenn die Kinder nicht nur gleich leistungsstark und motiviert waren, sondern von den Lehrkräften selbst auch als vergleichbar motiviert und anstrengungsbereit eingeschätzt wurden, scheinen für die sozialen Verzerrungen in Leistungserwartungen weitere Annahmen eine Rolle zu spielen. Interview- und MouseLab-Studien zeigten für Übergangsempfehlungen, dass Lehrkräfte neben schulischen Leistungen, Motivation und Arbeitsverhalten auch der Verfügbarkeit elterlicher Unterstützung Bedeutung beimessen (Böhmer, I., Hörstermann, Gräsel, Krolak-Schwerdt & Glock, 2015; Nölle, Hörstermann, Krolak-Schwerdt & Gräsel, 2009). Annahmen zur elterlichen Unterstützung könnten den verbleibenden Anteil sozialer Verzerrungen in Leistungserwartungen möglicherweise erklären. Entsprechend könnten Lehrkräfte annehmen, Kinder aus sozial bessergestellten Familien erhielten in schulischen Belangen mehr Unterstützung von ihren Eltern und würden daher auch höhere Leistungszuwächse erzielen als Kinder aus sozial schlechtergestellten Familien. Dieser Vermutung hat sich die Autorin der vorliegenden Arbeit mit weiterführenden Analysen anhand eines einzelnen Items zur Lehrkräfteeinschätzung der elterlichen Unterstützung angenähert (nicht in Teilstudie 2 berichtet). Sowohl für sprachliche als auch für mathematische Leistungserwartungen erwies sich diese Variable zusätzlich zu den Leistungen und den Selbst- und Fremdeinschätzungen der Motivation und des Arbeitsverhaltens als prädiktiv. Nach Hinzunahme der Lehrkräfteeinschätzung der elterlichen Unterstützung war der Effekt des sozioökonomischen Status auf die Leistungserwartungen in beiden Fächern nicht mehr signifikant.⁷ Diese Analysen liefern erste Hinweise darauf, dass mit dem sozialen Hintergrund assoziierte Verzerrungen von Leistungserwartungen nicht nur auf Annahmen der Lehrkräfte über die Motivation und das Arbeitsverhalten der Kinder, sondern auch auf Annahmen über die elterliche Unterstützung zurückgeführt werden können. Der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens der Kinder scheint dabei allerdings eine größere Bedeutung zuzukommen als den Annahmen zur elterlichen Unterstützung. Angaben der Eltern zur tatsächlichen Unterstützung konnten nicht

⁷ Für die ethnischen und geschlechtsbezogenen Verzerrungen in Leistungserwartungen änderten sich die Ergebnisse nicht. Die Lehrkraftannahmen über die elterliche Unterstützung scheinen diesen Analysen zufolge also für ethnische und geschlechtsbezogene Verzerrungen in Leistungserwartungen keine Rolle zu spielen.

berücksichtigt werden. Deshalb bleibt unklar, ob die Lehrkraftwahrnehmung zutrifft oder möglicherweise ebenfalls stereotype Annahmen widerspiegelt. Diese Analysen sind daher nur als erste Annäherung zu werten. Zukünftige Studien sollten der Frage explizit und mit optimierter Operationalisierung nachgehen. Zusätzlich dazu, dass eine Skala aus mehreren Items zur Lehrkrafteinschätzung der elterlichen Unterstützung anzustreben wäre, sollten auch Elternangaben berücksichtigt und als Vergleichsmaß für die Lehrkrafteinschätzungen herangezogen werden.

Geschlecht – Unterschiede und Verzerrungen in Erwartungen und Einschätzungen sowie Rückschlüsse auf Stereotype

Im Hinblick auf das Geschlecht der Schüler*innen zeigen die Ergebnisse der Teilstudien, dass Mädchen insgesamt als motivierter und anstrengungsbereiter von ihren Lehrkräften gesehen werden als Jungen. Zudem erwarten die Lehrkräfte für Mädchen einen höheren Leistungszuwachs im sprachlichen Bereich als für Jungen. Den Jungen sprechen sie hingegen einen Leistungsvorsprung in Mathematik zu. Während sich der antizipierte Leistungsvorsprung der Jungen auch im eingesetzten Mathematiktest zeigte, überschätzten die Erwartungen der Lehrkräfte im sprachlichen Bereich Unterschiede in den beobachteten Leistungen. Ebenso antizipierten die Lehrkräfte einen größeren Vorteil der Mädchen in ihrer Motivation und ihrem Arbeitsverhalten als Unterschiede in den Angaben der Mädchen und Jungen zur ihrer Lernfreude und Anstrengungsbereitschaft nahelegten.

Die Richtung der von den Lehrkräften angenommenen Leistungsunterschiede stimmt mit Befunden nationaler und internationaler Schulleistungsstudien für Viert- bzw. Neuntklässler*innen überein. Diese wiesen wiederkehrend auf einen Leistungsvorsprung der Mädchen im Lesen hin (z. B. Viertklässler*innen bei IGLU 2016: McElvany, Kessels, Schwabe & Kasper, 2017; 15-Jährige bei PISA 2009: Naumann, Artelt, Schneider & Stanat, 2010), während die Jungen höhere Kompetenzen in Mathematik erzielten (z. B. Viertklässler*innen bei TIMSS 2015: Wendt, Steinmayr & Kasper, 2016; 15-Jährige bei PISA 2012: Sälzer, Reiss, Schiepe-Tiska, Prenzel & Heinze, 2013). Der Vorteil der Mädchen im Lesen fiel allerdings deutlich größer aus als der Vorsprung der Jungen in Mathematik. Im IQB-Bildungstrend 2016 für Viertklässler*innen beispielsweise erreichten Mädchen im Lesen 24 Punkte mehr, wohingegen der Vorsprung der Jungen in Mathematik 19 Punkte betrug (Schipolowski, Wittig, Weirich & Böhme, 2017). Setzt man diese Unterschiede ins Verhältnis zu den in einer Normierungsstudie des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) identifizierten durchschnittlichen Lernzuwächsen von der dritten zur vierten Klasse (Behrens, Böhme & Krelle, 2009; Böhme & Bremerich-Vos, 2009; vgl. auch Schipolowski et al., 2017), entspricht der Vorsprung der Mädchen im Lesen etwa einem Drittel bis halben Schuljahr, während sich der Vorsprung der Jungen auf etwa ein Fünftel eines Schuljahrs beläuft. Auch dieser stärkere Vorteil von Mädchen im Lesen

gegenüber dem Vorteil von Jungen in Mathematik spiegelt sich akkurat in den Leistungserwartungen der Lehrkräfte wider.

Gleichzeitig ergibt sich durch die in der vorliegenden Studie vergleichbaren sprachlichen Leistungen der Mädchen und Jungen eine positive Verzerrung in den Leistungserwartungen im Fach Deutsch zugunsten der Mädchen. Dass die sprachlichen Leistungserwartungen enger mit Befunden großangelegter Schulleistungsstudien übereinstimmen als mit den Testleistungen und motivationalen Merkmalen der konkreten Schüler*innen der Klasse, könnte auf die Anwendung von Geschlechterstereotypen – also auf eine eher kategorienbasierte Informationsverarbeitung – hindeuten. Die Stereotype, wie sie in den Leistungserwartungen sichtbar geworden sind, könnten auf Basis der Ergebnisse der Schulleistungsstudien als verhältnismäßig akkurat eingestuft werden, wenn man das Alter der Schüler*innen ignoriert. Während sich TIMSS und IGLU auf Viertklässler*innen und die PISA-Studien auf Neuntklässler*innen beziehen, handelt es sich in der vorliegenden Stichprobe um Erstklässler*innen kurz nach der Einschulung. Entsprechend besteht die Frage, ob sich leistungsbezogene Geschlechterdisparitäten über die Bildungslaufbahn verändern. Blieben sie relativ stabil, wäre eine Übertragung des Wissens über Geschlechterunterschiede in höheren Klassenstufen auf niedrigere Klassenstufen unproblematisch und dürfte auf Gruppenebene zu akkurateren Erwartungen und Einschätzungen beitragen. Verändern sie sich allerdings bedeutsam, würde die Anwendung dieses Wissens in systematischen Fehleinschätzungen resultieren. Empirische Befunde deuten an, dass lesebezogene Fähigkeiten von Jungen und Mädchen in der frühen Grundschulzeit noch weitgehend ähnlich ausgeprägt sind, sich über die Grundschulzeit hinweg aber ein Vorteil der Mädchen herausbildet, der sich in der Sekundarstufe zunehmend verstärkt (vgl. Stanat et al., 2018). Im mathematischen Bereich deuten sich hingegen bereits im Vorschulalter Vorteile der Jungen gegenüber Mädchen an, die über die Schullaufbahn hinweg weitgehend stabil bleiben (vgl. Stanat et al., 2018). Auf dieser Grundlage könnte es durchaus sein, dass die Lehrkräfte in beiden Fächern stereotype Annahmen über Geschlechtergruppen, die auf medienwirksam berichteten Ergebnissen von Schulleistungsstudien basieren, angewendet haben. Da die beobachteten Geschlechterunterschiede in den Mathematikleistungen der Erstklässler*innen der vorliegenden Stichprobe mit denen in späteren Schuljahren weitgehend übereinstimmten, würde die Anwendung der Stereotypen (ebenso wie die Verarbeitung individueller Informationen) zu akkuraten Erwartungen beitragen. Im Fach Deutsch hingegen wurden ebenfalls Vorteile der Mädchen angenommen, die sich aber erst später in der Schullaufbahn zeigen, wodurch es zu den positiven Verzerrungen in den Erwartungen im Fach Deutsch für Mädchen gekommen sein könnte.

Möglicherweise könnte diese variierende Akkuratheit von Geschlechterstereotypen für verschiedene Altersgruppen bzw. Schulstufen ein Grund für den uneinheitlichen internationalen und nationalen Forschungsstand sein. Wenn geschlechterspezifische Verzerrungen identifiziert wurden,

fielen diese übereinstimmend mit Geschlechterstereotypen im sprachlichen Bereich und in fächerübergreifenden Maßen typischerweise zugunsten der Mädchen und in Mathematik typischerweise zugunsten der Jungen aus. Ein substantieller Anteil der Studien hat jedoch keine geschlechterspezifischen Verzerrungen identifiziert (vgl. Kapitel 2.2.4). Würde die obige These zutreffen, sollten Verzerrungen eher bei Erwartungen und Einschätzungen für jüngere Kinder beobachtet werden, auf die die Ergebnisse großangelegter Schulleistungsstudien für Viert- und Neuntklässler*innen nur bedingt übertragbar sind. Für die bisherigen Studien aus Deutschland könnte dies einem groben Vergleich zufolge zutreffen. Die beiden Studien, die keine Verzerrungen in Lehrkräfteeinschätzungen im Lesen und in Mathematik identifiziert hatten (Karing et al., 2011; Schrader & Helmke, 1990), untersuchten Schüler*innen am Ende der fünften bzw. sechsten Klasse; Stahl (2007) und Tiedemann (2000, 2002), deren Befunde auf geschlechterstereotype Verzerrungen im Lesen und in Mathematik hindeuten, analysierten hingegen Daten von Dritt- und Viertklässler*innen.

Des Weiteren deuten die Ergebnisse zu Geschlechterunterschieden in der Lehrkräfteeinschätzung zur Motivation und zum Arbeitsverhalten auf die Bedeutung eines Stereotyps hin, das Mädchen eine höhere Motivation und ein günstigeres Arbeitsverhalten zuspricht. Dieser Befund steht im Einklang mit bisherigen Studien, denen zufolge Lehrkräfte Mädchen im Vergleich zu Jungen als motivierter und anstrengungsbereiter ansehen sowie davon ausgehen, Mädchen würden ihre Hausaufgaben sorgfältiger erledigen und sich im Unterricht angemessener verhalten als Jungen (vgl. Kapitel 2.2.5). Über bisherige Forschung hinausgehend zeigt die vorliegende Studie, dass sich diese antizipierten Vorteile der Mädchen nur zu einem geringen Ausmaß in den Selbsteinschätzungen der Lernenden widerspiegeln. Möglicherweise spielt für die günstigeren Einschätzungen von Mädchen das Stereotyp des angepassten, fleißigen und sorgfältigen Mädchens eine Rolle; beziehungsweise umgekehrt für die verhältnismäßig ungünstige Einschätzung von Jungen das Stereotyp des ungezogenen, faulen und unsorgfältigen Jungen. Es könnte aber auch sein, dass Unterschiede im Unterrichtsverhalten der Mädchen und Jungen, die sich nicht in den Selbsteinschätzungen der Kinder widerspiegeln, für die Geschlechterunterschiede in den Einschätzungen der Lehrkräfte verantwortlich sind (siehe Kapitel 5.2.2).

Die Analysen der Teilstudie 2 verdeutlichen weiterhin, dass die Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens eine wichtige Rolle für die Untersuchung von geschlechterbezogenen Verzerrungen in Leistungserwartungen spielen. Auf der Grundlage theoretischer Überlegungen sollten Leistungserwartungen für Jungen und Mädchen umso stärker übereinstimmen, je ähnlicher sich die Lernenden hinsichtlich individueller Merkmale sind (Robinson-Cimpian et al., 2014). Für beobachtbare Geschlechterunterschiede in den Leistungserwartungen trifft dies jedoch nur zum Teil zu. Werden gleich leistungsstarke Jungen und Mädchen von ihren

Lehrkräften auch als ähnlich motiviert und anstrengungsbereit eingeschätzt, erwarten die Lehrkräfte zwar vergleichbare Leistungsentwicklungen für Mädchen und Jungen im Lesen. In Mathematik verdeutlichen die Ergebnisse jedoch, dass Lehrkräfte bei gleich leistungsstarken Jungen und Mädchen, die sie als ähnlich motiviert und anstrengungsbereit einschätzen, einen größeren Lernzuwachs für die Jungen erwarten. Dieses Ergebnis ist mit der Annahme vereinbar, Lehrkräfte würden ihre eigentlich höheren Erwartungen an Jungen in Mathematik nach unten adjustieren, da sie sie als weniger motiviert und anstrengungsbereit einschätzen.

5.1.2 Effekte von Leistungserwartungen und deren Prozesse

Der zweite Schwerpunkt der vorliegenden Arbeit bezog sich auf Zusammenhänge von Leistungserwartungen mit der Leistungsentwicklung von Erstklässler*innen und deren Vermittlungsprozesse. Ausgehend von bisherigen Studien zu selbsterfüllenden Prophezeiungen im Schulkontext wurde angenommen, dass sich inakkurat hohe bzw. niedrige Erwartungen in höheren bzw. niedrigeren Lernzuwächsen bei den Schüler*innen niederschlagen und diese Effekte durch differentielles Lehrkraftverhalten – spezifisch dem Feedback der Lehrkräfte – zum Teil vermittelt werden. Die Ergebnisse der Teilstudie 3 belegen das Auftreten von Erwartungseffekten sowohl für das Fach Deutsch als auch für das Fach Mathematik. Ebenfalls waren zwei der drei untersuchten Feedbackmerkmale mit inakkuraten Erwartungen assoziiert. Eine bedeutsame Mediation der Erwartungseffekte durch das Lehrkraftfeedback ließ sich jedoch nicht nachweisen. Im Folgenden sollen zunächst die Ergebnisse zu Zusammenhängen inakkurater Erwartungen mit der Leistungsentwicklung genauer dargestellt und basierend auf dem bisherigen Forschungsstand diskutiert werden. Daran anschließend wird der Blick auf die Befunde zur Mediation gerichtet.

Befunde zum Zusammenhang inakkurater Erwartungen und Leistungsentwicklungen

Die Befunde belegen das Auftreten von Erwartungseffekten in ersten Grundschulklassen in Deutschland für beide untersuchten fachlichen Domänen. Im sprachlichen Bereich fielen die Effekte dabei stärker aus als im mathematischen Bereich ($\beta = .21$ versus $\beta = .09$). Ebenso zeigten sich im sprachlichen Bereich sowohl positive als auch negative Effekte von Leistungserwartungen, während in Mathematik lediglich positive Effekte inakkurat hoher Erwartungen auftraten. Für die unterschiedliche Stärke der Erwartungseffekte im sprachlichen und mathematischen Bereich kann es verschiedene Gründe geben. Eine mögliche Erklärung könnte sein, dass Leistungsindikatoren im sprachlichen Bereich mehr Raum für Interpretation bieten, während sich Leistungen im mathematischen Bereich klarer beurteilen lassen. Dies könnte sich in einer geringeren Akkuratheit

von Urteilen im sprachlichen als im mathematischen Bereich niederschlagen und entsprechend die Wahrscheinlichkeit von Erwartungseffekten erhöhen. Gegen diese Annahme spricht allerdings der Befund, dass die Anteile akkurater Erwartungen in beiden Domänen relativ ähnlich ausfielen (Deutsch: 34-37 %, Mathematik: 36-37 %; siehe dazu auch in Kapitel 5.1.1 den Abschnitt zum Ausmaß der (In)Akkuratheit von Leistungserwartungen). Gleichzeitig deuten Befunde zur diagnostischen Kompetenz zwar darauf hin, dass es individuellen Lehrkräften oft in einer fachlichen Domäne besser gelingt zu akkuraten Urteilen zu gelangen als in einer anderen fachlichen Domäne (Lorenz, C. & Artelt, 2009). Ein Vorteil eines spezifischen Fachs mit Blick auf die Urteilsgüte wurde aber nicht gefunden (z. B. Südkamp et al., 2012; vgl. auch Kapitel 2.2.3). Eine weitere mögliche Erklärung steht im Zusammenhang mit dem Vorwissen der Kinder zum Zeitpunkt der Einschulung und daraus resultierenden unterschiedlichen Lernzuwächsen. Während viele Kinder zum Zeitpunkt der Einschulung bereits bis 10 oder 20 zählen und einfache Additions- und Subtraktionsaufgaben erfolgreich lösen (vgl. Deutscher & Selter, 2013), können vergleichsweise wenige neu eingeschulte Kinder bereits Buchstaben oder Wörter lesen (vgl. Juska-Bacher, 2013). Im ersten Grundschuljahr könnte Lehrkräften demnach bei der Vermittlung grundlegender Lesefähigkeiten eine besonders starke Bedeutung zukommen, wodurch ihr Einfluss und der ihrer Erwartungen auf den Lernzuwachs im Lesen stärker ausfällt als in Mathematik.

Die Stärke der identifizierten Erwartungseffekte korrespondiert eng mit den in Reviews berichteten Ergebnissen von $r = .10$ bis $r = .20$ (z. B. Jussim et al., 2009). Den Richtlinien von Cohen (1988) zufolge handelt es sich um kleine Effektstärken. Basierend auf Richtlinien zu Effektstärken bei der Erforschung individueller Unterschiede, die auf 708 Metaanalysen zurückgehen und entsprechend empirisch fundiert sind, können die gefundenen Zusammenhänge als relativ klein bis durchschnittlich eingestuft werden (Gignac & Szodorai, 2016). Zur Frage, ob positive oder negative Effekte von Erwartungen überwiegen, zeigen die Ergebnisse der vorliegenden Dissertation kein klares Bild und könnten auf domänenspezifische Unterschiede hindeuten. Im sprachlichen Bereich zeigten sich sowohl positive („Galatea-Effekt“) als auch negative Erwartungseffekte („Golem-Effekt“), wobei der Leistungsnachteil unterschätzter Schüler*innen stärker ausfiel als der Vorteil überschätzter Schüler*innen. In Mathematik traten hingegen nur positive Erwartungseffekte auf. Lernende, denen die Lehrkräfte inakkurat niedrige Erwartungen entgegenbrachten, unterschieden sich in ihrem Lernzuwachs in Mathematik nicht von Lernenden, denen die Lehrkräfte akkurate Erwartungen entgegenbrachten. Dass im Fach Mathematik vor allem positive Erwartungseffekte zu verzeichnen sind, steht im Einklang mit den Ergebnissen von Madon et al. (1997). Spezifische Befunde zu Unterschieden in positiven und negativen Erwartungseffekten für den sprachlichen Bereich lagen bislang nicht vor (vgl. Kapitel 2.3.2). Die vorliegende Dissertation erweitert den Forschungsstand diesbezüglich.

Befunde zum Prozess von Erwartungseffekten: Zusammenhänge zwischen Erwartungen, Feedback und Leistungsentwicklung

Zur Analyse der Prozesse von Erwartungseffekten zeigten die Ergebnisse von Teilstudie 3 zunächst, dass zwei der drei untersuchten Feedbackmerkmale mit der Höhe der inakkuraten Erwartungen variierten. So erhielten Kinder, denen die Lehrkräfte inakkurat höhere Erwartungen entgegenbrachten, mehr Feedback zu ihren Leistungen als zu ihrem Verhalten. Das Leistungsfeedback war für diese Kinder außerdem zu einem höheren Anteil positiv. Zu welchem Anteil die Lehrkraft den Lernenden im Falle eines Fehlers weitere Hinweise gab, variierte hingegen nicht in Abhängigkeit der inakkuraten Erwartungen. Die Stärke des Zusammenhangs zwischen inakkuraten Erwartungen und den beiden Feedbackmerkmalen lag im kleinen Bereich ($\beta = .10$ bis $.13$). Etwa 3 bzw. 4 % des Feedbacks, das ein*e Schüler*in erhielt, fiel günstiger aus, wenn die inakkuraten Erwartungen um einen Skalenwert höher ausfielen. Damit unterstützen die Befunde die Annahme, dass Leistungserwartungen teilweise über das Feedback, das Lehrkräfte ihren Schüler*innen geben, kommuniziert werden. Dieses Ergebnis steht im Einklang mit Befunden von Studien, die sich mit einer Risikogruppe von Lernenden befassten (Montague & Rinaldi, 2001) bzw. Angaben der Schüler*innen über die Feedbackpraxis ihrer Lehrkräfte heranzogen (Brattesani et al., 1984; Chen et al., 2011). Über diese Befunde bisheriger Studien hinausgehend zeigt die vorliegende Studie, dass Erwartungen von Lehrkräften und ihr Feedbackverhalten auch nach Kontrolle beobachteter Leistungen, allgemeiner kognitiver Fähigkeiten und motivationaler Merkmale der Lernenden miteinander assoziiert sind.

Weiterführende Analysen zur Mediation des Zusammenhangs zwischen Leistungserwartungen und Leistungsentwicklungen der Schüler*innen ergaben allerdings keine Hinweise darauf, dass das Feedback den Zusammenhang bedeutsam mediiert. Dies dürfte primär darin begründet liegen, dass die untersuchten Feedbackmerkmale die Leistungsentwicklung der Schüler*innen – mit Ausnahme des Anteils positiven Leistungsfeedbacks im Fach Mathematik – nicht bedeutsam vorhersagten, wenn die Ausgangsleistungen der Schüler*innen, ihre allgemeinen kognitiven Fähigkeiten, ihre Motivation und ihr familiärer Hintergrund kontrolliert wurden. Hierfür sind verschiedene Gründe denkbar. Erstens wäre es möglich, dass bestehende Zusammenhänge aufgrund des relativ kleinen Ausschnitts der Lehrkraft-Kind-Interaktionen, der untersucht wurde, nicht nachgewiesen werden konnten. Die Videoanalysen der vorliegenden Studie basierten je Klasse auf zwei bis vier Unterrichtsstunden in der Mitte des ersten Grundschuljahrs. Obgleich dieser Umfang an Beobachtungsmaterial für Videostudien bereits recht groß ist, bildet dieser nur einen Bruchteil der gesamten Unterrichtsinteraktionen ab, die über ein Schuljahr hinweg zwischen den Lehrkräften und Schüler*innen stattfinden. Gleichzeitig erscheint plausibel, dass sich Unterrichtsprozesse über ein Schuljahr hinweg verändern und die Videokodierungen daher nur eine Momentaufnahme der

spezifisch untersuchten Unterrichtsstunden bilden. Solche Veränderungsprozesse könnten in der ersten Zeit nach der Einschulung besonders stark ausgeprägt sein, da sich Lehrkräfte und Kinder zunächst kennenlernen, schulische Verhaltensregeln etabliert und Unterrichtsroutinen geschaffen werden müssen.

Zweitens könnten die ausbleibenden Zusammenhänge zwischen Feedback und Leistungsentwicklung auf die Herausforderung zurückgehen, qualitativ hochwertiges Feedback detailliert zu erfassen. Anhand drei verschiedener Aspekte wurde in der vorliegenden Studie versucht, lernrelevante Merkmale individuellen Feedbacks abzubilden. Basierend auf theoretischen Überlegungen und Befunden aus der Lehr-Lern-Forschung (für genauere Informationen siehe Theorieteil der Teilstudie 3) wurde das Verhältnis von Leistungs- und Verhaltensfeedback sowie von positivem und negativem Leistungsfeedback untersucht und ob im Fall von negativem Leistungsfeedback weitere Hinweise gegeben wurden. Dass Zusammenhänge zwischen diesen Feedbackaspekten und der Leistungsentwicklung ausgeblieben sind, könnte allerdings als Hinweis darauf gewertet werden, dass zukünftige Videokodierungen noch weiter in die Tiefe gehen müssten. Beispielsweise bleibt auf der Grundlage des in der vorliegenden Studie genutzten Kodierschemas unklar, ob die Lehrkraft positives Leistungsfeedback auf der Ebene der Performanz in einer spezifischen Aufgabe (z. B. „Super, du hast alle Wörter fehlerfrei vorgelesen, auch die schweren.“) oder auf der Ebene der*des Lernenden als Person (z. B. „Super, du bist eine tolle Vorleserin.“) formuliert. Die erstgenannte Art von Feedback ist Studien zufolge deutlich effektiver für den Lernerfolg von Schüler*innen (z. B. Hattie & Timperley, 2007). Detailliertere Kodierungen, die solche und ähnliche Aspekte qualitativ hochwertigen Feedbacks differenzierter erfassen, wären wünschenswert, waren in der KuL-Studie aufgrund begrenzter zeitlicher und finanzieller Ressourcen aber nicht möglich. Wie solche detaillierteren Videokodierungen in zukünftigen Forschungsprojekten aussehen könnten, wird in Kapitel 5.3 skizziert.

Ein dritter Erklärungsansatz bezieht sich auf Merkmale der Schüler*innen, die den Zusammenhang zwischen Feedback und Leistungsentwicklung vermitteln oder verändern. Das detailliertere Modell von Jussim (1986) zum Prozess der Erwartungseffekte geht davon aus, dass psychologische Merkmale der Schüler*innen als Bindeglied zwischen Lehrkraftverhalten und leistungsbezogenen Reaktionen der Schüler*innen wirken. Etwa könnte individuelles Feedback zunächst auf das Selbstkonzept der Lernenden oder ihre Motivation wirken, was sich in unterschiedlich intensiven Lernanstrengungen niederschlägt und darüber letztlich zu verschiedenen hohen Lernzuwächsen beiträgt. Würde es sich um eine reine Mediation handeln, sollten sich auch ohne Berücksichtigung der motivationalen Merkmale zum Schuljahresende korrelative Zusammenhänge zwischen den Leistungen und dem individuellen Feedback zeigen. Wahrscheinlicher als eine reine Vermittlung ist allerdings, dass psychologische Prozesse auf Seiten der Lernenden die

Wirkungsweise des Feedbacks mitbestimmen und verändern. Diese Annahme steht im Einklang mit dem Verständnis von Unterricht als Angebotsstruktur für erfolgreiche Lernprozesse, deren Nutzung und Wirkungsweise von Bedingungen und Prozessen auf Seiten der Schüler*innen abhängt (Angebots-Nutzungs-Modell; Helmke, 2017; Seidel, 2014). Demzufolge kann das gleiche positive Feedback einer Lehrkraft von den Lernenden unterschiedlich interpretiert werden, etwa basierend auf vorherigen Unterrichtserfahrungen oder aufgrund von Überzeugungen über die eigene Leistungsfähigkeit. So könnte eine positive Äußerung der Lehrkraft zum korrekten Lösen einer relativ einfachen Aufgabe von den Lernenden als ehrliche Bestätigung wahrgenommen oder als implizite Abwertung der eigenen Leistungsfähigkeit interpretiert werden, was sich wiederum unterschiedlich auf Lernanstrengungen auswirken kann. Entsprechend ist es möglich, dass sich in der vorliegenden Studie gegenläufige Prozesse für verschiedene Gruppen von Schüler*innen überlagert haben, weshalb im Mittel kein Zusammenhang zwischen individuellem Feedback und Leistungsentwicklung nachgewiesen werden konnte. Für beide Annahmen, also dass psychologische Merkmale auf Seiten der Schüler*innen einerseits als Moderator (z. B. Madon et al., 1997) und andererseits als Mediator (z. B. Friedrich et al., 2015; Trusz, 2018) wirken können, liegen erste empirische Hinweise vor. Die Befundlage ist allerdings noch relativ überschaubar und die Ergebnisse variieren teilweise auch innerhalb von Studien je nach untersuchter abhängiger Variable. Beispielsweise zeigten Befunde von Friedrich et al. (2015), dass die Effekte von Leistungserwartungen der Lehrkräfte auf die Mathematiknoten am Ende der fünften Klasse, aber nicht auf Leistungen in einem Mathematiktest, teilweise vom akademischen Selbstkonzept der Lernenden mediiert wurden.

Die genannten Erklärungsansätze könnten dafür verantwortlich sein, dass in der vorliegenden Studie kein klarer Zusammenhang zwischen Feedback und Leistungsentwicklung und damit auch keine bedeutsame Mediation der Erwartungseffekte durch das Feedback nachgewiesen werden konnte. Wenn diese Annahmen (teilweise) zutreffen, wäre es möglich, dass Feedback entgegen der Befunde der vorliegenden Dissertation Erwartungseffekte bedeutsam mediiert. Gleichzeitig ist das Befundmuster der vorliegenden Arbeit erstaunlich konsistent zu den aggregierten Ergebnissen der Metaanalyse von Harris und Rosenthal (1985), in der der gemittelte Zusammenhang zwischen Erwartungen und unterschiedlichen Feedbackmerkmalen $r = .13$ und zwischen Feedbackmerkmalen und Schüler*innenoutcomes $r = .07$ betrug. Die Schlussfolgerungen der Metaanalyse, dass (1) das Feedback von Lehrkräften zwar mit ihren Erwartungen variiert, aber (2) kein starker Mediator von Erwartungseffekten auf die Leistungen von Schüler*innen ist, lässt sich auch auf der Grundlage der vorliegenden Ergebnisse halten. Zu (1) hat Teilstudie 3 das Wissen hinzugefügt, dass die Assoziation zwischen Erwartungen und Feedback über beobachtbare Unterschiede in lernrelevanten Ressourcen der Schüler*innen hinausgeht, also auch *inakkurat* hohe bzw. niedrige Erwartungen mit unterschiedlichem Feedback assoziiert sind. Hinsichtlich der Vermittlungsprozesse von

Erwartungseffekten bedeutet dies allerdings, dass weitere Analysen notwendig sind. Die vorliegende Dissertation konnte nur einen sehr kleinen Ausschnitt potenziell relevanten Lehrkräfteverhaltens berücksichtigen und war auch in Bezug auf die untersuchten Merkmale des Feedbacks eingeschränkt. Entsprechend konnte mit dieser Arbeit nur ein kleines Puzzleteil zum genaueren Verständnis der Entstehung von Erwartungseffekten beigetragen werden.

5.2 Grenzen der Arbeit

Die vorliegende Arbeit erweitert den Forschungsstand zu Erwartungen von Lehrkräften und ihren Auswirkungen auf die Lernentwicklung von Schüler*innen auf der Basis eines längsschnittlich angelegten Forschungsprojekts. Die erhobenen Daten zeichnen sich durch einige methodische Vorteile aus, die Limitationen bisheriger Studien überwinden. Dies betrifft zunächst die Erfassung der Erwartungen und Einschätzungen der Lehrkräfte sowie der Ausgangsfähigkeiten der Schüler*innen zu einem Zeitpunkt, an dem sich Kinder und Lehrkräfte erst seit wenigen Wochen kannten. Die Wahrscheinlichkeit, dass bis zur Datenerhebung Erwartungen der Lehrkräfte bereits die Fähigkeiten der Kinder über den Prozess einer selbsterfüllenden Prophezeiung beeinflusst haben, dürfte daher relativ gering sein. Gleichzeitig reduziert dieses Vorgehen das Risiko, dass die Erwartungen zu einem bedeutsamen Anteil auf Merkmalen der Lernenden basieren, die in dieser Studie nicht erfasst wurden (siehe unten). Eine weitere Besonderheit der Studie ist die Untersuchung von Unterrichtsprozessen anhand von Videografien. Die videobasierten Auswertungen der Lehrkraft-Kind-Interaktionen ermöglichen es, die Entstehung von Erwartungseffekten zu erforschen anstatt lediglich zu untersuchen, ob solche Zusammenhänge auftreten oder nicht.

Trotz dieser wichtigen Vorteile gegenüber bisheriger Forschung weist auch die vorliegende Studie Limitationen auf. Die Einschränkungen der einzelnen Teilstudien wurden bereits in den jeweiligen Publikationen detailliert diskutiert. An dieser Stelle sollen die zentralen Grenzen aufgegriffen werden, die alle Teilstudien und damit die Arbeit im Gesamten betreffen. Die Hauptlimitationen beziehen sich auf die Generalisierbarkeit der Befunde, methodische Einschränkungen sowie Erkenntnisse über Kausalität.

5.2.1 Generalisierbarkeit der Befunde

Alle Teilstudien basieren auf Daten der KuL-Studie, die im Schuljahr 2013/2014 in ersten Grundschulklassen in Nordrhein-Westfalen durchgeführt wurde. Entsprechend ist die Stichprobe nicht repräsentativ für Deutschland und die Ergebnisse somit nur eingeschränkt verallgemeinerbar.

Hinzu kommt, dass die Teilnehmenden aufgrund der Freiwilligkeit der Studienteilnahme selektiert sein können. Auf der Ebene der Lehrkräfte kann angenommen werden, dass sich besonders engagierte Personen zur Studienteilnahme entschieden haben. Geht man davon aus, dass engagierte Lehrkräfte besonders bemüht sind, die Kinder ihrer Klasse gut kennenzulernen und ihre lernbezogenen Ausgangslagen genau einzuschätzen, sollte dies zu einer Überschätzung der Akkuratheit von Erwartungen und Einschätzungen in der Studie verglichen mit der Akkuratheit von Urteilen der Gesamtheit der Grundschullehrkräfte führen. Dies hätte für die Analysen der Teilstudien 1 und 2 zur Konsequenz, dass sowohl die inakkuraten Anteile der Leistungserwartungen als auch deren Verzerrungen unterschätzt wurden. Daraus resultierend könnten in Teilstudie 3 auch die Zusammenhänge zwischen inakkuraten Erwartungen und Feedbackverhalten bzw. Leistungsentwicklung unterschätzt worden sein.

Auch auf Ebene der Schüler*innen könnte die freiwillige Teilnahme zu Selektionen beigetragen haben. Gemessen am durchschnittlichen *Highest International Socio-Economic Index of Occupational Status* (HISEI) von $M = 49.00$, der im IQB-Bildungstrend 2016 für Viertklässler*innen in Nordrhein-Westfalen ermittelt wurde (Haag et al., 2017), ist die KuL-Stichprobe hinsichtlich des sozioökonomischen Status der Kinder (HISEI: $M = 52.51$) leicht positiv selektiert. Für die Analysen der Teilstudien 1 und 2 könnte hieraus resultieren, dass die sozialen Verzerrungen in den Leistungserwartungen von Lehrkräften etwas unterschätzt wurden, da Kinder mit besonders niedrigerem sozialen Status, denen besonders niedrige Erwartungen entgegengebracht werden dürften, in der Stichprobe unterrepräsentiert sind. Gleiches trifft möglicherweise auch auf Verzerrungen in Abhängigkeit vom ethnischen Hintergrund der Schüler*innen zu. Zum einen waren auch Kinder, von denen mindestens ein Elternteil im Ausland geboren ist, zu einem geringeren Anteil in der KuL-Stichprobe repräsentiert (36 %) als an Grundschulen in Nordrhein-Westfalen allgemein (40 %; Rjosk et al., 2017). Zum anderen wurden Kinder der ersten Zuwanderungsgeneration von den Analysen der Teilstudien 1 und 2 ausgeschlossen, da diese Gruppe von Erstklässler*innen erst relativ kurz in Deutschland lebt, was Erwartungen und Einschätzungen von Lehrkräften in besonderer Weise beeinflussen könnte.

Zudem basieren die Analysen zum Prozess von Erwartungseffekten (Teilstudie 3) auf einer relativ kleinen Stichprobe, was die Belastbarkeit und Verallgemeinerbarkeit der Ergebnisse zusätzlich einschränkt. Hervorzuheben ist jedoch, dass sich die an der Videostudie teilnehmenden Lehrkräfte und Schüler*innen hinsichtlich vielfältiger Merkmale nicht bedeutsam von der Gesamtstichprobe der KuL-Studie unterscheiden. Die Ergebnisse der einzelnen Teilstudien basieren demnach auf einer hinsichtlich zentraler Variablen vergleichbaren Datengrundlage.

5.2.2 Methodische Einschränkungen

Auch auf methodischer Ebene bestehen Limitationen, die alle Teilstudien betreffen. Eine Einschränkung ist mit dem Residualansatz (Madon et al., 1997) verbunden, der genutzt wurde, um den inakkuraten Anteil in den Leistungserwartungen der Lehrkräfte zu bestimmen. Dieser Ansatz geht davon aus, dass jegliche Variation in Leistungserwartungen, die nicht von den als valide Indikatoren definierten Variablen erklärt wird, inakkurate Anteile kennzeichnet. Dieses Vorgehen ist mit der Problematik verbunden, dass die Inakkuratheit erstens aufgrund von Unterschieden in unbeobachteten Merkmalen der Schüler*innen und zweitens aufgrund von Messfehlern in den einbezogenen Prädiktoren überschätzt worden sein könnte. Ersteres wäre dann der Fall, wenn Lehrkräfte Unterschiede in lernrelevanten Merkmalen bei den Lernenden wahrnehmen und in ihre Erwartungen einbeziehen, die nicht in den Analysen berücksichtigt wurden. In diesem Fall würde unerklärte Varianz in den Leistungserwartungen keine Inakkuratheit widerspiegeln, sondern auf unbeobachtete akkurate Einflüsse zurückgehen. Zwar lässt sich diese Möglichkeit nicht gänzlich ausschließen. Das breitere Verständnis von Akkuratheit, das den Teilstudien zugrunde gelegt wurde, sollte das Risiko aber verringern; als valide Indikatoren von Leistungserwartungen wurden nicht nur fachspezifische Leistungen, sondern auch allgemeine kognitive Fähigkeiten und motivationale Merkmale der Lernenden berücksichtigt. Zusätzlich dürfte das Design der KuL-Studie die Problematik abmildern. Zum Zeitpunkt der Erfassung von Erwartungen und lernrelevanten Merkmalen der Schüler*innen kannten sich die Lehrkräfte und Schüler*innen erst wenige Wochen. Dadurch ist es nicht nur unwahrscheinlich, dass die Erwartungen der Lehrkräfte die Leistungen der Schüler*innen vor der Erhebung bereits beeinflusst hatten, sondern auch, dass die Erwartungen und Einschätzungen der Lehrkräfte bereits maßgeblich von individuellen Merkmalen der Lernenden geprägt worden sind, die in der vorliegenden Studie nicht erfasst wurden.

Auch die Problematik möglicher Messfehler lässt sich nicht vollständig ausschließen. Anders als die Herausforderung *unbeobachteter* Merkmale von Schüler*innen bezieht sich diese Problematik auf die *beobachteten* Merkmale der Schüler*innen. Auch die besten Erhebungsinstrumente stellen nur eine näherungsweise Schätzung der „wahren“ Leistungen und Fähigkeiten der Schüler*innen dar. Solange allerdings der erhobene Testwert zufällig um den „wahren Wert“ streut, sollten die in den Teilstudien beobachteten Zusammenhänge erwartungstreu sein, in ihrer Stärke aber eher unterschätzt werden.

Die Problematik von unbeobachteten Merkmalen der Schüler*innen und von Messfehlern betreffen auch die Analysen zur Einschätzung der Motivation und des Arbeitsverhaltens in Teilstudie 2. In der vorliegenden Arbeit konnten lediglich Selbsteinschätzungen der Lernfreude und der Anstrengungsbereitschaft der Lernenden als Vergleichsmaß für die Einschätzungen der Lehrkräfte

einbezogenen werden. Dieses Vorgehen ist in der Forschung zur diagnostischen Kompetenz bei Urteilen über motivationale Merkmale gängig (vgl. Praetorius & Südkamp, 2019) und basiert auf der Annahme, dass die betreffende Person selbst am besten in der Lage ist, Auskunft über ihren innerlichen, motivationalen Zustand zu geben (z. B. Givvin et al., 2001; vgl. auch Kapitel 2.2.5). Gleichzeitig muss sich die innerliche Motivation nicht unbedingt in konsistentem beobachtbarem Verhalten niederschlagen, wodurch akkurate Einschätzungen dieser Merkmale erschwert werden (Givvin et al., 2001; Praetorius et al., 2015; Urhahne & Zhu, 2015). Dass ein Kind im Unterricht beispielweise zurückhaltend ist und sich nur selten am Unterrichtsgeschehen beteiligt, könnte sowohl Ausdruck mangelnder Motivation als auch anderer sozioemotionaler Merkmale wie etwa sozialer Unsicherheit oder Ängstlichkeit sein. Um zu Einschätzungen über die motivationalen Merkmale von Schüler*innen zu gelangen, müssen sich Lehrkräfte aber letztlich auf solches beobachtbares Unterrichtsverhalten der Lernenden stützen. Für die Ergebnisse der vorliegenden Arbeit ist es daher möglich, dass beobachtbare Unterschiede in den Lehrkräfteeinschätzungen anstatt auf inakkurate Stereotype auf tatsächlich beobachtbare Unterschiede im Unterrichtsverhalten der Kinder zurückzuführen sind, die sich nicht in den Selbsteinschätzungen der Kinder auf den beiden Skalen Lernfreude und Anstrengungsbereitschaft widerspiegeln. Ob und zu welchem Anteil dies der Fall ist, sollten zukünftige Studien untersuchen, indem sie neben Selbsteinschätzungen der Lernenden auch objektive Beobachtungen des Unterrichtsverhaltens der Schüler*innen in ihre Untersuchungen einbeziehen.

5.2.3 Ökologische Validität und Erkenntnisse über Kausalität

Weiterhin bewegt sich auch die vorliegende Studie im Spannungsfeld zwischen ökologischer Validität und Erkenntnissen über Kausalität. Die vorliegende Dissertation basiert auf nicht-experimentellen Daten von realen Lehrkräften und Schüler*innen in Klassenverbänden. Auf dieser Grundlage können die generierten Erkenntnisse als ökologisch valide eingestuft werden. Gleichzeitig ist dieses Vorgehen – stärker als experimentelle Untersuchungen – mit dem Risiko möglicher Alternativerklärungen verbunden.

Durch den querschnittlichen Charakter der Teilstudien 1 und 2 ist zudem bei der Interpretation der Wirkrichtung der untersuchten Zusammenhänge Vorsicht geboten. Während in Teilstudie 1 die Annahme unplausibel ist, Leistungserwartungen von Lehrkräften hätten sich in den soziodemografischen Merkmalen der Schüler*innen niedergeschlagen, lässt sich in Teilstudie 2 nicht ausschließen, dass die Leistungserwartungen möglicherweise auch die Einschätzungen der Motivation und des Arbeitsverhaltens beeinflusst haben. Durch die längsschnittliche Anlage und das

Erhebungsdesign sind die Ergebnisse von Teilstudie 3 von diesem Problem weniger betroffen. Da sowohl die Ausgangsfähigkeiten der Schüler*innen als auch die Leistungserwartungen der Lehrkräfte kurz nach Beginn des Lehrkraft-Kind-Kontakts erhoben wurden, können Merkmale der späteren Lehrkraft-Kind-Interaktion, wie z. B. das Feedback, diese Maße nicht vorab beeinflusst haben. Statistische Zusammenhänge zwischen den Erwartungen zu Schuljahresbeginn und dem Feedbackverhalten der Lehrkräfte in der Mitte des Schuljahrs können daher entsprechend der zeitlichen Reihung nur als Pfad von den Erwartungen zum Feedback interpretiert werden und nicht umgekehrt. Gleiches gilt für die Zusammenhänge zwischen Erwartungen bzw. Feedback und Leistung der Schüler*innen am Schuljahresende.

Für alle drei Teilstudien gilt allerdings, dass die identifizierten Zusammenhänge keine direkten kausalen Beziehungen zwischen den Lehrkraft- und Schüler*innenvariablen darstellen müssen. Vielmehr könnten (und dürften) zwischengelagerte Prozesse eine Rolle spielen, die in der vorliegenden Dissertation nicht untersucht werden konnten. Bei den Teilstudien 1 und 2 wird beispielweise aufgrund theoretischer Überlegungen angenommen, dass der ethnische Hintergrund, der soziale Hintergrund und das Geschlecht der Schüler*innen von Lehrkräften wahrgenommen und dadurch stereotype Annahmen über die Gruppen aktiviert werden. Aufgrund der Ergebnisse zu Verzerrungen in Leistungserwartungen und Einschätzungen der Motivation und des Arbeitsverhaltens wird angenommen, dass eine kategorienbasierte Informationsverarbeitung stattgefunden hat und Stereotype über Gruppen angewendet wurden. Die genaue Prozessabfolge wurde allerdings nicht untersucht, sodass darüber keine empirisch fundierte Aussage getroffen werden kann. Ebenso könnte es in Teilstudie 3 sein, dass sich die Erwartungen der Lehrkräfte nicht direkt im Feedback niederschlagen, sondern zunächst psychologische Prozesse auf Seiten der Lehrkräfte auftreten (siehe Kapitel 2.3.1). Gleiches ist auch für den Zusammenhang zwischen inakkuraten Erwartungen und Leistungsentwicklungen denkbar.

5.3 Perspektiven für die weitere Forschung

Aus den Ergebnissen der vorliegenden Dissertation ergeben sich vielfältige Perspektiven für zukünftige Forschung. An dieser Stelle sollen zwei übergreifende Themen für mögliche Anschlussvorhaben beispielhaft skizziert werden: (1) Unterschiede zwischen Lehrkräfte und Klassen, (2) Weiterführende Analysen zum Prozess von Erwartungseffekten.

5.3.1 Unterschiede zwischen Lehrkräften bzw. Klassen

Die Ergebnisse der vorliegenden Dissertation deuten auf einen bedeutsamen inakkuraten Anteil von Leistungserwartungen, auf Verzerrungen für verschiedene Gruppen von Schüler*innen und auf Zusammenhänge mit dem Feedbackverhalten der Lehrkräfte und der Leistungsentwicklung der Lernenden hin. Zwar wurde in allen Analysen die Mehrebenenstruktur der Daten (Schüler*innen geclustert in Klassen bzw. Lehrkräften) berücksichtigt, etwaige Unterschiede zwischen Lehrkräften und Klassen im Auftreten und Ausmaß der (In)Akkuratheit, Verzerrungen und Erwartungseffekte war jedoch nicht Gegenstand der Analysen. Dabei ist es durchaus plausibel, dass verschiedene Lehrkräfte zu unterschiedlich akkuraten Urteilen gelangen und sich verschieden stark von soziodemografischen Merkmalen der Schüler*innen leiten lassen. Ebenso ist denkbar, dass kontextuelle Merkmale der Klasse und ihrer Zusammensetzung das Ausmaß der (In)Akkuratheit und Verzerrungen beeinflussen und das Auftreten von Erwartungseffekten fördern oder hemmen. Wie in den Kapiteln 2.2.3 und 2.3.2 ausgeführt, existieren verschiedene Annahmen, welche Merkmale dabei eine Rolle spielen können. Die empirische Befundlage ist zu den meisten Annahmen allerdings uneinheitlich bzw. stützt diese nur unzureichend. Weitere Forschung, die etwaige Unterschiede zwischen Lehrkräften und Klassen systematisch untersucht und mehrere individuelle und kontextuelle Merkmale simultan einbezieht, könnte zu einem besseren Verständnis beitragen.

Anstatt dabei die (In)Akkuratheit von Erwartungen zu fokussieren, wäre es aufgrund der identifizierten Vorteile hoher Erwartungen (näheres dazu in Kapitel 5.4.2) vielversprechend, zu untersuchen, welche Lehrkräfte von sich aus hohe Erwartungen an alle Schüler*innen richten und was andere Lehrkräfte darin unterstützen kann, solche hohen Erwartungen auszubilden. So könnte es möglicherweise Lehrkräften leichter fallen, hohe Erwartungen für ihre Schüler*innen zu entwickeln, wenn sie generell mit ihrem Beruf als Lehrkraft zufrieden sind und eine hohe Selbstwirksamkeitserwartung als Lehrkraft haben (Burić & Kim, 2020; Tschannen-Moran & Hoy, 2001). Ebenso könnte die implizite Theorie, dass Leistungen und Intelligenz veränderbar sind, hohe Erwartungen begünstigen.

Für ein fundiertes Verständnis der Entstehung von Verzerrungen sollten zukünftige Studien Stereotype von Lehrkräften explizit einbeziehen. Beispielsweise dürften sich Lehrkräfte, die ethnische, soziale oder geschlechtsbezogene Stereotype stärker teilen, eher von diesen verallgemeinernden Annahmen bei der Urteilsbildung leiten lassen, was (im Falle von inakkuraten Stereotypen) stärkere Verzerrungen der Erwartungen und Einschätzungen zur Folge hätte. Aktuelle Studien belegt dies für ethnische (Lorenz, G., 2019) und geschlechtsbezogene (Muntoni & Retelsdorf, 2018) Leistungsstereotype und Leistungserwartungen von Lehrkräften. Ein genaueres Verständnis solcher

und ähnlicher Prozesse ist etwa für die Aus- und Fortbildung von Lehrkräften relevant, wenn es darum geht, den Einfluss stereotyper Annahmen zu reduzieren.

Auch Erwartungseffekte könnten in Klassen von Lehrkräften, die Stereotype stärker teilen, besonders stark ausfallen. Ähnlich zu den Ergebnissen der Arbeitsgruppe um Babad (Babad, 2009; Babad et al., 1989a) zu *high-bias* und *low-bias* Lehrkräften (vgl. Kapitel 2.3.2) könnte angenommen werden, dass es Lehrkräften, die Stereotype stärker teilen, weniger gut gelingt, Ungleichbehandlungen beim Unterrichten von Schüler*innen der betreffenden Gruppen zu vermeiden. Demnach könnten Erwartungseffekte in Klassen, deren Lehrkräfte Stereotype stärker teilen, größer ausfallen als in Klassen, deren Lehrkräfte keine oder nur gering ausgeprägte Stereotype über verschiedene Gruppen von Schüler*innen haben. Gleichermaßen wäre es hierbei aufschlussreich zu untersuchen, ob Erwartungen für verschiedene Schüler*innen auch in Klassen von Lehrkräften, die an alle Schüler*innen hohe Erwartungen richten, (auf einem höheren Gesamtniveau) variieren und ob auch in diesen Klassen individuell unterschiedliche Lernzuwächse in Abhängigkeit von der Höhe der Erwartungen nachweisbar sind.

5.3.2 Weiterführende Analysen zum Prozess von Erwartungseffekten

Trotz des Einbezugs drei verschiedener Merkmale von Feedback konnte die vorliegende Dissertation nur ein kleines Puzzleteil zum Verständnis der Prozesse von Erwartungseffekten beitragen. Zum genaueren Verständnis, welche Verhaltensweisen von Lehrkräften Erwartungseffekte vermitteln, bedarf es weiterer Forschung. Zukünftige Vorhaben sollten dabei einerseits vertiefte Analysen qualitativ hochwertigen Feedbacks durchführen, andererseits aber auch weitere Verhaltensweisen untersuchen, durch die sich Erwartungen auf die Leistungsentwicklung auswirken können.

Um qualitativ hochwertiges Feedback detailliert abzubilden, wäre es zunächst sinnvoll, die Lehrkraft-Kind-Interaktionen wortwörtlich zu transkribieren. Auf dieser Grundlage lassen sich spezifischere Kodierungen des Gesagten vornehmen. Den Arbeiten von Hattie und Timperley (2007) zufolge, beinhaltet qualitativ hochwertiges Feedback Informationen über (1) das angestrebte Lernziel, (2) den bereits erzielten Lernfortschritt und (3) die nächsten notwendigen Schritte, um das Lernziel zu erreichen. Dabei werden vier Ebenen unterschieden, auf die sich das Feedback beziehen kann: (1) Ebene der Aufgabe, (2) Ebene des Lernprozesses, (3) Ebene der Selbstregulation, (4) Ebene der Person. Feedback auf der Ebene der Person gilt dabei verglichen mit den drei anderen Ebenen als weniger effektiv. Bei detaillierten Feedbackkodierungen wäre anzustreben, all diese Informationen abzubilden. Zudem würden genaue Informationen zur Angemessenheit und Qualität der Äußerung

bzw. des Verhaltens der*des Lernenden, die dem Feedback voranging, einen zusätzlichen Erkenntnisgewinn liefern.

Zusätzlich zu einer detaillierteren Kodierung des Feedbacks sind weitere Verhaltensweisen von Lehrkräften von Interesse, die an der Entstehung von Erwartungseffekten beteiligt sein können. Hierbei ist zunächst an Verhaltensweisen zu denken, die sich den drei weiteren von Harris und Rosenthal (1985) identifizierten Dimensionen Klima, Input und Output zuordnen lassen (vgl. Kapitel 2.3.1 und 2.3.3). In natürlich auftretenden Unterrichtssituationen scheinen der Metaanalyse folgend die beiden Aspekte positives und negatives Klima, also allgemeine positive bzw. negative Einstellungen, Äußerungen oder Verhaltensweisen der Lehrkraft gegenüber dem*der Schüler*in, besonders bedeutsam zu sein. Für diese beiden Aspekte des Klimas der Lehrkraft-Kind-Interaktion fand die Metaanalyse in nicht-experimentellen Studien einen Zusammenhang von $r = .26$ (positives Klima) bzw. $r = .30$ (negatives Klima) mit den Erwartungen von Lehrkräften. Da beide Aspekte zudem relativ stark mit den Schüler*innenoutcomes zusammenhingen (positives Klima: $r = .40$, negatives Klima $r = .36$; Ergebnisse experimenteller und nicht-experimenteller Studien wurden nicht unterschieden), könnten sie einen substantiellen Anteil von Erwartungseffekten, die im alltäglichen Unterrichtsgeschehen auftreten, mediiieren. Zur Erhebung des Klimas von Lehrkraft-Kind-Interaktionen können dabei verschiedene Perspektiven herangezogen werden – die Perspektive der Lehrkraft, des Kindes oder einer beobachtenden, an der Interaktion unbeteiligten Person. Zur Untersuchung von Erwartungseffekten sollte zumindest die Perspektive unbeteiligter Beobachter*innen berücksichtigt werden. Dies liegt zunächst darin begründet, dass Beobachtungsdaten objektivere Einschätzungen darstellen und differenzierter vorgenommen werden können. Zudem zeigte eine Metaanalyse von Cornelius-White (2007), dass sich die kognitive und emotional-motivationale Entwicklung von Schüler*innen am stärksten von Merkmalen der Lehrkraft-Kind-Interaktion vorhersagen ließ, wenn sie auf Beobachtungsdaten basierten. Die Perspektiven der Lehrkräfte und Schüler*innen waren zwar ebenfalls substantiell, aber zu einem deutlich geringeren Ausmaß mit den Outcomes der Schüler*innen assoziiert. Aus diesen Gründen sollte der Beitrag des Klimas der Lehrkraft-Kind-Interaktion zu Erwartungseffekte in jedem Fall basierend auf Beobachtungsdaten untersucht werden. Ein multiperspektivischer Ansatz, in den auch die Sichtweisen der Lehrkräfte und Schüler*innen einbezogene werden, könnte jedoch zusätzliche Erkenntnisse liefern.

Doch auch Lehrkraftverhaltensweisen, die sich nicht direkt den vier Dimensionen Klima, Input, Output und Feedback zuordnen lassen, könnten für die Vermittlung von Erwartungseffekten bedeutsam sein (vgl. Jussim et al., 2009). Zukünftige Forschung sollte daher auch weitere, bislang nicht untersuchte Verhaltensweisen in den Blick nehmen. Vielversprechend könnte es hierzu sein, Erkenntnisse zur allgemeinen Unterrichtsqualität (z. B. Helmke, 2017; Meyer, 2004) in die Forschung

zur Vermittlung von Erwartungseffekten einzubeziehen. Zum Teil überschneiden sich die in diesen Kontexten diskutierten Unterrichtspraktiken und Verhaltensweisen mit den vier bereits untersuchten Dimensionen von Lehrkraftverhalten, teilweise liefern sie aber auch spezifischere Erkenntnisse zu relevanten Aspekten oder gehen über die vier Dimensionen hinaus. Jussim et al. (2009) beziehen sich beispielsweise bei ihren Überlegungen zu weiteren, möglicherweise relevanten Verhaltensweisen von Lehrkräften auf Aspekte der Unterrichtsqualität, die im Kontext der Motivationsförderung unter dem Begriff des *TARGET framework* diskutiert werden (z. B. Lüftenegger, van de Schoot, Schober, Finsterwald & Spiel, 2014). TARGET steht dabei für *Task, Authority/Autonomy, Recognition, Grouping, Evaluation* und *Time* (Blumenfeld, 1992; Epstein, 1988). Jussim et al. (2009) leiten aus dem TARGET-Ansatz verschiedene Unterrichtspraktiken ab, die in der Forschung zu Erwartungseffekten noch selten untersucht wurden, aber durchaus relevant sein könnten. Dies betrifft beispielsweise die Gestaltung von Aufgaben. Die Input-Dimension von Rosenthal bezog sich vorwiegend auf die Menge und den Schwierigkeitsgrad von Aufgaben. Darüber hinaus können die Aufgaben, die Lehrkräfte ihren Schüler*innen geben, aber auch in Abhängigkeit von ihren Erwartungen darin variieren, wie ansprechend und bedeutungsvoll die Aufgaben gestaltet sind (entsprechend der *Task*-Dimension von TARGET, die sich auf eine ansprechende Gestaltung von Aufgaben bezieht, die heraus- aber nicht überfordern). Auch die *Time*-Dimension von TARGET bietet Hinweise für weitere lohnenswerte Untersuchungsgegenstände bei der Vermittlung von Erwartungseffekten. So wurde bislang vor allem untersucht, ob sich die absolute Interaktionsdauer und -häufigkeit mit den Lernenden in Abhängigkeit von den Erwartungen von Lehrkräften unterscheidet (Harris & Rosenthal, 1985). Relevanter für die Motivations- und Leistungsentwicklung dürfte allerdings sein, zu welchem Anteil die gemeinsame Zeit zur Erschließung von Lerninhalten genutzt und ob sie kognitiv anregend gestaltet wird. Im Kontext von TARGET werden zudem heterogene Lerngruppen, die kooperativ an Aufgaben arbeiten, als motivationsförderlich erachtet (*Grouping*). Die Einteilung der Lerngruppen könnte allerdings auch von den Erwartungen der Lehrkraft bestimmt werden, was sich ungünstig auf die Motivation derjenigen Kinder auswirken könnte, die einer leistungsschwachen Gruppe zugewiesen werden. Aus der Perspektive von TARGET könnten solche aus unterschiedlichen Erwartungen resultierenden Ungleichbehandlungen dazu führen, dass die Motivation und das Interesse von Schüler*innen, denen die Lehrkräfte hohe Erwartungen entgegenbringen, gestärkt wird, was zu umfangreicheren Lernanstrengungen dieser Lernenden beitragen und dadurch letztlich zu höheren Lernzuwächsen führen dürfte.

Diese angenommene Wirkkette verdeutlicht einen weiteren Aspekt, der in zukünftiger Forschung zu den Prozessen von Erwartungseffekten verstärkt berücksichtigt werden sollte. Für viele Lehrkraftverhaltensweisen wird angenommen, dass sie sich zunächst auf psychologische Merkmale von Lernenden, wie etwa ihre Motivation, ihr Interesse, ihr Selbstkonzept oder ihre

Selbstwirksamkeitserwartung auswirken (vgl. Kapitel 2.3.1). Bei einer tiefergehenden Erforschung, wie Erwartungseffekte entstehen, wäre es daher wichtig, psychologische Merkmale von Schüler*innen als Mediatoren aber auch als mögliche Moderatoren einzubeziehen.

5.4 Bedeutung der Erkenntnisse für die pädagogische Praxis

Die Thematik der vorliegenden Studie ist von hoher Relevanz für die pädagogische Praxis. Im Folgenden sollen die Befunde der Arbeit zunächst hinsichtlich ihrer Bedeutung für den Beitrag von Erwartungseffekten zu Bildungsungleichheiten diskutiert werden. Daran anschließend werden Implikationen für die Aus- und Weiterbildung von Lehrkräften abgeleitet.

5.4.1 Erwartungen von Lehrkräften und Bildungsungleichheiten

Eine Motivation für die vorliegende Dissertation liegt in den mittlerweile gut belegten ethnischen, sozialen und geschlechterbezogenen Bildungsungleichheiten begründet. Wie in den Teilstudien gezeigt wurde, variieren Leistungserwartungen von Lehrkräften schon in der ersten Grundschulklasse systematisch in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund und Geschlecht der Schüler*innen. Darüber hinaus erfüllen sich inakurate Erwartungen zum Teil selbst, sodass Kinder abhängig von den Erwartungen der Lehrkräfte höhere oder niedrigere Lernzuwächse im ersten Grundschuljahr erzielen. Da Kinder mit einem niedrigeren sozialen Status allgemein und Kinder mit einem türkischen Zuwanderungshintergrund sowie Jungen im Lesen häufiger unterschätzt werden, sind diese Kinder mit einer höheren Wahrscheinlichkeit von negativen Erwartungseffekten betroffen. Zusammengefasst legen die Befunde der vorliegenden Arbeit daher nahe, dass Leistungserwartungen von Lehrkräften zur Entstehung oder Verfestigung von ethnischen, sozialen und geschlechterbezogenen Leistungsdisparitäten beitragen und sich letztlich auch in Ungleichheiten bei Noten, Bildungsübergängen und dem Zertifikatserwerb niederschlagen können. Wenngleich diese kumulative Evidenz für einen Beitrag von Erwartungseffekten zu Bildungsungleichheiten spricht, sollten empirische Studien dieser These explizit nachgehen und den etwaigen Erklärungsbeitrag quantifizieren. In den letzten Jahren haben Forschungsarbeiten erste Hinweise zu dieser Frage generiert. Vor allem zu sozialen Leistungsdisparitäten scheinen Erwartungseffekte bedeutsam beizutragen (z. B. Hollenstein et al., 2019; Ready & Chu, 2015). Die höheren Erwartungen an Kinder aus sozial bessergestellten Familien könnten demnach dazu beitragen, dass die Lehrkräfte mit ihnen eine bessere Beziehung aufbauen und die Interaktionen lernförderlicher gestalten als für Kinder aus sozial benachteiligten Familien, was sich letztlich in höheren Lernzuwächsen niederschlagen dürfte.

Auch für zuwanderungs- und geschlechtsbezogene Disparitäten weisen erste Studien auf eine mögliche Bedeutung von Erwartungseffekten hin, die Befundlage ist aber weniger eindeutig (Zuwanderungshintergrund: z. B. Lorenz, G., 2018; Ready & Chu, 2015; Geschlecht: z. B. Gentrup & Rjosk, 2018; Heyder, Steinmayr & Kessels, 2019; Muntoni & Retelsdorf, 2018).

5.4.2 Aus- und Weiterbildung – Streben nach hohen Erwartungen für alle Schüler*innen

Eine wichtige Implikation für die pädagogische Praxis ergibt sich aus dem Befund, dass hohe Erwartungen sowohl im Lesen als auch in Mathematik mit besonders günstigen Lernentwicklungen bei den Schüler*innen einhergingen. Dabei erwiesen sich inakkurat hohe Erwartungen gemessen am Lernzuwachs der Kinder auch als günstiger als akkurate Erwartungen. Demzufolge sollten Lehrkräfte dazu ermutigt werden, hohe Erwartungen an alle Schüler*innen ihrer Klasse zu richten, selbst wenn die Erwartungen auf der Grundlage der Ausgangsfähigkeiten und Motivation der Lernenden möglicherweise zu optimistisch wirken. Vor diesem Hintergrund ist auch der Befund, dass etwa 60 % der Unterschiede in Leistungserwartungen nicht durch beobachtbare Unterschiede in den Leistungen, allgemeinen kognitiven Fähigkeiten und motivationalen Merkmalen von Schüler*innen erklärt werden können, weniger besorgniserregend. Für den individuellen Lernprozess von Schüler*innen scheint weniger eine Rolle zu spielen, wie akkurat die Erwartungen lernrelevante Merkmale widerspiegeln, als vielmehr, wie positiv die Erwartungen ausgeprägt sind. Dies stellt jedoch nicht zwangsweise die pädagogische Relevanz diagnostischer Kompetenz von Lehrkräften in Frage. Aktuelle Studien deuten darauf hin, dass es Lehrkräften mit hoher diagnostischer Kompetenz besser gelingt, eine hohe Unterrichtsqualität zu realisieren (Hill & Chin, 2018; Westphal, Gronostaj, Vock, Emmrich & Harych, 2016). Gleichzeitig scheinen gute diagnostische Fähigkeiten für den effektiven Einsatz von Unterrichtstechniken – etwa von Leistungsfeedback (Behrmann & Souvignier, 2013) oder Strukturierungs- und Unterstützungsmaßnahmen (Helmke & Schrader, 1987; Schrader & Helmke, 1987) – wichtig zu sein. Möglicherweise kommt diagnostischen Fähigkeiten von Lehrkräften vor allem für die Unterrichtsgestaltung auf Klassenebene Bedeutung zu, während für individuelle Lehrkraft-Kind-Interaktionen positive Werthaltungen – wie etwa hohe Erwartungen – vorteilhaft sind.

Trotz einer Ermutigung zu hohen Erwartungen, dürfte es allerdings unrealistisch sein, dass Lehrkräfte für alle Schüler*innen exakt gleichhohe Erwartungen ausbilden. Auch bei der Implementation hoher Erwartungen an alle Schüler*innen besteht daher für Lehrkräfte die Herausforderung, Einflüsse von soziodemografischen Merkmalen der Schüler*innen und darauf basierenden Stereotypen zu reduzieren. Wie also können Lehrkräfte in Aus- und Weiterbildung darin unterstützt werden, stereotypen Einflüssen auf ihre Erwartungen und Einschätzungen

entgegenzuwirken? Sozialpsychologische Experimente verdeutlichen, dass es möglich ist, die Auswirkungen von Stereotypen abzumildern (für einen Überblick siehe Blair, 2002). Das Streben danach, Stereotype zu unterdrücken, erwies sich allerdings als eher unverlässliche Strategie, die teilweise die Wirkung von Stereotypen sogar verstärkte (z. B. Galinsky & Moskowitz, 2000). Vielversprechender scheint das Aktivieren von sogenannten Gegenstereotypen (*counterstereotypes*). Blair, Ma und Lenton (2001) beispielsweise baten die Studienteilnehmer*innen sich etwa fünf Minuten lang eine starke Frau (Gegenstereotyp) vorzustellen bevor sie verschiedene Tests bearbeiteten, die das Ausmaß der Aktivierung von Stereotypen erfassen. Diejenigen Teilnehmer*innen, die sich eine starke Frau vorgestellt hatten, zeigten bedeutsam seltener stereotypen-konformes Verhalten in den Tests als die Teilnehmenden aller anderen Experimentalbedingungen, u. a. derjenigen Gruppe, die das Stereotyp einer schwachen Frau versuchen sollte zu unterdrücken. Ein anderes Beispiel, das sich leichter auf die Aus- und Weiterbildung von Lehrkräften übertragen lässt, lieferten Rudman, Ashmore und Gary (2001). Sie zeigten in einem quasi-experimentellen Studiendesign, dass sich automatische Stereotype und Vorurteile von Studierenden, die an einem einsemestrigen Kurs zum Thema „Vorurteile und Konflikt“ teilnahmen, über das Semester hinweg stärker reduzierten als in der Vergleichsgruppe. Ein erster Schritt in Richtung Reduktion stereotyper Einflüsse kann es demnach sein, Module zum Thema „Heterogenität in Schule und Unterricht“ in den Studienordnungen zu etablieren und in einzelnen Veranstaltungen für Stereotype und deren mögliche Auswirkungen zu sensibilisieren und Gegenstrategien einzuüben. Wie die Ergebnisse der vorliegenden Dissertation nahelegen, beziehen sich schulisch relevante Stereotype nicht nur auf Leistungen von Schüler*innen sondern auch auf ihre Motivation und ihr Arbeitsverhalten. Daher sollten bei der Sensibilisierung für Stereotype und bei der Einübung von Gegenstrategien auch verallgemeinernde Überzeugungen über andere lernrelevante Merkmale thematisiert werden. Zu berücksichtigen bleibt dabei ferner, dass ein Merkmal von Lernenden, das als akkurater Indikator für zukünftige Leistungen gelten kann – in diesem Fall die Motivation und das Arbeitsverhaltens –, von den Lehrkräften ebenfalls zum Teil inakkurat und in Abhängigkeit von soziodemografischen Merkmalen der Lernenden eingeschätzt wird, was wiederum zu Verzerrungen im Leistungsurteil führen kann. Interventionen, die Stereotype und ihre Auswirkungen thematisieren, dürften in der Ausbildung von Lehramtsstudierenden insgesamt noch vielversprechender sein als spätere Fortbildungen mit erfahrenen Lehrkräften, bei denen sich stereotype Erwartungsmuster und Urteilsprozesse bereits verfestigt haben können (vgl. Glock & Krolak-Schwerdt, 2013).

Da Stereotype u. a. auf gesellschaftlich geteiltem Wissen basieren, kommt einem sensiblen Umgang mit Aussagen über Gruppen aber auch in anderen Lehrveranstaltungen des Lehramtsstudiums Bedeutung zu, etwa wenn Befunde zu Leistungsdisparitäten aus großangelegten

Schulleistungsstudien vorgestellt werden. Um Übergeneralisierungen vorzubeugen, sollten die Befunde mit den Studierenden kritisch diskutiert und auch anhand der Variation *innerhalb* von Gruppen und den Überschneidungsbereichen zwischen Gruppen eingeordnet werden.

Um die Vorteile hoher Erwartungen zu nutzen, wäre es darüber hinaus wichtig, angehende und praktizierende Lehrkräfte nicht nur darüber zu informieren, dass sich Erwartungen generell selbst erfüllen können, sondern insbesondere auch auf die positiven Effekte hoher Erwartungen hinzuweisen. Dies eröffnet die Frage, wie Aus- und Weiterbildung Lehrkräfte darin unterstützen kann, hohe Erwartungen an alle Schüler*innen zu entwickeln und die positiven Effekte zu nutzen. Zur Etablierung hoher Erwartungen an alle Schüler*innen liegen erste vielversprechende Befunde von Interventionsstudien vor (Übersicht und Metaanalyse von 19 Interventionen: de Boer, Timmermans & van der Werf, 2018). Im Mittel war es sowohl möglich, die Erwartungen der Lehrkräfte zu erhöhen, als auch die Leistungen von Schüler*innen dadurch positiv zu beeinflussen. Sowohl die Ansätze als auch die spezifischen Befunde variierten jedoch stark. Besonders häufig wurden Interventionen identifiziert, die auf eine Verhaltensänderung bei den Lehrkräften abzielten und / oder eine generelle Bewusstheit dafür schaffen wollten, dass Erwartungen die Leistungsentwicklungen von Schüler*innen beeinflussen können und / oder dass Erwartungen inakkurat und / oder dass Erwartungen verzerrt sein können. Weiterhin adressierten einige Interventionen direkt die Überzeugungen der Lehrkräfte zu Leistungen von Schüler*innen. Eine beispielhafte Intervention, die auf Bewusstheitsschaffung und Verhaltensänderungen abzielte, entwickelten Rubie-Davies, Peterson, Sibley und Rosenthal (2015). Sie schulten die an der Intervention teilnehmenden Lehrkräfte darin, Unterrichtspraktiken einzusetzen, die Lehrkräfte mit hohen Erwartungen an alle Schüler*innen typischerweise anwenden (z. B. an Lernzielen ausgerichtetes Feedback; Rubie-Davies, Hattie, Townsend & Hamilton, 2007). Über einen Zeitraum von zwei Monaten fanden insgesamt vier Workshops statt, die ergänzt wurden durch Praxisphasen, in denen die Teilnehmenden das Gelernte in ihren Unterrichtsalltag implementierten. Außerdem fanden videobasierte Coachings und Reflexionstreffen mit den Wissenschaftler*innen statt. Über den Interventionszeitraum zeigten Schüler*innen, die von den Lehrkräften der Experimentalgruppe unterrichtet wurden, in Mathematik einen höheren Lernzuwachs als Schüler*innen, deren Lehrkräfte zur Kontrollgruppe gehörten. Im Lesen war dieser Vorteil der Experimentalgruppe nicht zu beobachten. Interviews mit den Lehrkräften zufolge lag die ausbleibende Wirkung im Lesen darin begründet, dass sie die gelernten Praktiken vorwiegend im Mathematikunterricht etabliert hatten, da es ihnen dort leichter fiel als im Sprachunterricht.

Solche und ähnliche Interventionen können vielversprechende Ansätze sein, um für das Auftreten von Erwartungseffekten zu sensibilisieren und die positiven Effekte hoher Erwartungen zu nutzen. Gleichzeitig bleibt in weiterer Forschung noch zu klären, ob solche Interventionen zur

Etablierung hoher Erwartungen an alle Schüler*innen tatsächlich auch die Wahrscheinlichkeit für (zumindest negative) Erwartungseffekte reduzieren.

Literatur

- Alexander, K. & Schofield, J. W. (2006). Expectancy effects: How teachers' expectations influence student achievement. In J. W. Schofield (Hrsg.), *Migration Background, Minority-Group Membership and Academic Achievement Research Evidence from Social, Educational, and Developmental Psychology*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung gGmbH FSP Zivilgesellschaft, Konflikte und Demokratie Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration -AKI-.
- Alvidrez, J. & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology*, 91(4), 731–746. doi:10.1037//0022-0663.91.4.731
- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57(3), 175–193.
- Anders, Y., McElvany, N. & Baumert, J. (2010). Die Einschätzung lernrelevanter Schülermerkmale zum Zeitpunkt des Übergangs von der Grundschule auf die weiterführende Schule: Wie differenziert urteilen Lehrkräfte? In K. Maaz, J. Baumert, C. Gresch & N. McElvany (Hrsg.), *Der Übergang von der Grundschule in die weiterführende Schule: Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (S. 313–330). Bonn: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung.
- Anderson-Clark, T. N., Green, R. J. & Henley, T. B. (2008). The relationship between first names and teacher expectations for achievement motivation. *Journal of Language and Social Psychology*, 27(1), 94–99. doi:10.1177/0261927X07309514
- Archambault, I., Janosz, M. & Chouinard, R. (2012). Teacher beliefs as predictors of adolescents' cognitive engagement and achievement in mathematics. *The Journal of Educational Research*, 105(5), 319–328. doi:10.1080/00220671.2011.629694
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz. Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–137). Opladen: Leske + Budrich.
- Babad, E. (1990). Measuring and changing teachers' differential behavior as perceived by students and teachers. *Journal of Educational Psychology*, 82(4), 683–690. doi:10.1037/0022-0663.82.4.683
- Babad, E. (1993). Teachers' differential behavior. *Educational Psychology Review*, 5(4), 347–376.

- Babad, E. (1998). Preferential affect: The crux of the teacher expectancy issue. In J. Brophy (Hrsg.), *Advances in research on teaching: Expectations in the classroom* (Bd. 7, S. 183–214). Greenwich, CT: JAI Press.
- Babad, E. (2009). *The social psychology of the classroom*. New York: Taylor & Francis.
- Babad, E., Bernieri, F. & Rosenthal, R. (1989a). Nonverbal communication and leakage in the behavior of biased and unbiased teachers. *Journal of Personality and Social Psychology*, 56(1), 89–94.
- Babad, E., Bernieri, F. & Rosenthal, R. (1989b). When less information is more informative: Diagnosing teacher expectancies from brief samples of behaviour. *British Journal of Educational Psychology*, 59, 281–295. doi:10.1111/j.2044-8279.1989.tb03103.x
- Babad, E., Bernieri, F. & Rosenthal, R. (1991). Students as judges of teachers' verbal and nonverbal behavior. *American Educational Research Journal*, 28(1), 211–234. doi:10.3102/00028312028001211
- Babad, E., Inbar, J. & Rosenthal, R. (1982). Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology*, 74 (4), 469–474.
- Babad, E. & Taylor, P. J. (1992). Transparency of teacher expectancies across language, cultural boundaries. *The Journal of Educational Research*, 86(2), 120–125. doi:10.1080/00220671.1992.9941148
- Back, M. D. & Nestler, S. (2016). Accuracy of judging personality. In J. A. Hall, M. Schmid Mast & T. V. West (Hrsg.), *The social psychology of perceiving others accurately*. Cambridge: Cambridge University Press.
- Bates, C. & Nettelbeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, 21(2), 177–187. doi:10.1080/01443410020043878
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, J., Watermann, R. & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6(1), 46–71. doi:10.1007/s11618-003-0004-7
- Becker, D. (2012). *Pygmalion's long shadow. Determinants and outcomes of teacher evaluations*. Dissertation, Universität zu Köln, Köln.
- Becker, D. (2013). The impact of teachers' expectations on students' educational opportunities in the life course: An empirical test of a subjective expected utility explanation. *Rationality and Society*, 25(4), 422–469. doi:10.1177/1043463113504448

- Becker, D. & Birkelbach, K. (2013). Lehrer als Gatekeeper? Eine theoriegeleitete Annäherung an Determinanten und Folgen prognostischer Lehrerurteile. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen* (S. 207–237). Wiesbaden: Springer VS. doi:10.1007/978-3-531-18985-7
- Begeny, J. C., Eckert, T. L., Montarello, S. a. & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43–55. doi:10.1037/1045-3830.23.1.43
- Begeny, J. C., Krouse, H. E., Brown, K. G. & Mann, C. M. (2011). Teacher judgments of students' early literacy skills measured by the early literacy skills assessment: Comparisons of teachers with and without assessment administration experience. *Psychology in the Schools*, 47(8), 859–868.
- Behrens, U., Böhme, K. & Krelle, M. (2009). Zuhören – Operationalisierung und fachdidaktische Implikationen. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walter (Hrsg.), *Bildungsstandards Deutsch und Mathematik* (S. 357–375). Weinheim: Beltz.
- Behrmann, L. & Souvignier, E. (2013). The relation between teachers' diagnostic sensitivity, their instructional activities, and their students' achievement gains in reading. *Zeitschrift für Pädagogische Psychologie*, 27(4), 283–293. doi:10.1024/1010-0652/a000112
- Bennett, R. E., Gottesman, R. L., Rock, D. A. & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, 85(2), 347–356.
- Berscheid, E., Graziano, W., Monson, T. & Dermer, M. (1976). Outcome dependency: Attention, attribution, and attraction. *Journal of Personality and Social Psychology*, 34(5), 978–989. doi:10.1037/0022-3514.34.5.978
- Biggs, M. (2009). Self-fulfilling Prophecies. In P. Hedström & P. Bearman (Hrsg.), *The Oxford Handbook of analytic sociology* (S. 294–314). Oxford: Oxford University Press.
- Birkelbach, K. (2011). Lehrerurteile im Lebensverlauf. Valide Prognose oder Self-Fulfilling Prophecy? *Soziale Welt*, 62(3), 227–249.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6(3), 242–261.
- Blair, I. V., Ma, J. & Lenton, A. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of personality and social psychology*, 81(5), 828–841. doi:10.1037//0022-3514.81.5.828

- Blumenfeld, P. C. (1992). Classroom learning and motivation: Clarifying and expending goal theory. *Journal of Educational Psychology*, 84(3), 272–281.
- Bodenhausen, G. V., Kang, S. K. & Peery, D. (2012). Social categorization and the perception of social groups. In S. T. Fiske & C. N. Macrae (Hrsg.), *The SAGE Handbook of Social Cognition*: SAGE Publications.
- Böhme, K. & Bremerich-Vos, A. (2009). Lesekompetenzdiagnostik – Die Entwicklung eines Kompetenzstufenmodells für den Bereich Lesen. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walter (Hrsg.), *Bildungsstandards Deutsch und Mathematik* (S. 219–249). Weinheim: Beltz.
- Böhmer, I., Hörstermann, T., Gräsel, C., Krolak-Schwerdt, S. & Glock, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung: Welcher Urteilsregel folgen Lehrkräfte? *Journal für Bildungsforschung Online*, 7(2), 59–81.
- Böhmer, M., Englich, B. & Böhmer, I. (2017). Schülerbeurteilungen aus der Perspektive dualer Prozessmodelle der sozialen Urteilsbildung. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 50–54). Münster, New York: Waxmann.
- Brattesani, K. a., Weinstein, R. S. & Marshall, H. H. (1984). Student perceptions of differential teacher treatment as moderators of teacher expectation effects. *Journal of Educational Psychology*, 76(2), 236–247. doi:10.1037//0022-0663.76.2.236
- Brophy, J. E. & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. New York, NY: Holt, Rinehardt and Winston Inc.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.
- Burić, I. & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. *Learning and Instruction*, 66. doi:10.1016/j.learninstruc.2019.101302
- Chen, Y.-H., Thompson, M. S., Kromrey, J. D. & Chang, G. H. (2011). Relations of student perceptions of teacher oral feedback with teacher expectancies and student self-concept. *The Journal of Experimental Education*, 79(4), 452–477. doi:10.1080/00220973.2010.547888
- Chow, S. L. (1987). Some reflections on Harris and Rosenthal's 31 meta-analyses. *The Journal of Psychology*, 121(1), 95–100. doi:10.1080/00223980.1987.9712648
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78(2), 141–146.

- Cooper, H. M. (1979). Pygmalion grows up: A model for teacher expectation communication and performance influence. *Review of Educational Research*, 49(3), 389–410.
- Cooper, H. M. & Good, T. L. (1983). *Pygmalion grows up: Studies in the expectation communication process*. New York: Longman.
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77(1), 113–143. doi:10.3102/003465430298563
- Darley, J. M. & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, 35(10), 867–881.
- de Boer, H., Bosker, R. J. & van der Werf, M. P. C. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, 102(1), 168–179. doi:10.1037/a0017289
- de Boer, H., Timmermans, A. C. & van der Werf, M. P. C. (2018). The effects of teacher expectation interventions on teachers' expectations and student achievement: Narrative review and meta-analysis. *Educational Research and Evaluation*, 24(3–5), 180–200. doi:10.1080/13803611.2018.1550834
- de Boer, H. & van der Werf, M. P. C. (2015). Influence of misaligned parents' aspirations on long-term student academic performance. *Educational Research and Evaluation*, 21(3), 232–257. doi:10.1080/13803611.2015.1039548
- Demaray, M. K. & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, 13(1), 8–24.
- Deutscher, T. & Selter, C. (2013). Frühe mathematische Bildung – Forschungsbefunde und Förderkonzepte. In M. Stamm & D. Edelmann (Hrsg.), *Handbuch frühkindliche Bildungsforschung* (S. 543–556). Wiesbaden: Springer VS.
- Dicke, A.-L., Lüdtke, O., Trautwein, U., Nagy, G. & Nagy, N. (2012). Judging students' achievement goal orientations: Are teacher ratings accurate? *Learning & Individual Differences*, 22(6), 844–849. doi:10.1016/j.lindif.2012.04.004
- Diehl, C. & Fick, P. (2016). Ethnische Diskriminierung im deutschen Bildungssystem. In C. Diehl, C. Hunkler & C. Kristen (Hrsg.), *Ethnische Ungleichheiten im Bildungsverlauf* (S. 243–286). Wiesbaden: Springer VS.
- Doherty, J. & Conolly, M. (1985). How accurately can primary school teachers predict the scores of their pupils in standardised tests of attainment? A study of some non-cognitive factors that influence specific judgements. *Educational Studies*, 11(1), 41–60. doi:10.1080/0305569850110105

- Dompnier, B., Pansu, P. & Bressoux, P. (2006). An integrative model of scholastic judgments: Pupils' characteristics, class context, halo effect and internal attributions. *European Journal of Psychology of Education*, 21(2), 119–133.
- Duckworth, A. L. & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98(1), 198–208.
- Dusek, J. B. & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology*, 75(3), 327–346. doi:10.1037//0022-0663.75.3.327
- Dvir, T., Eden, D. & Banjo, M. L. (1995). Self-fulfilling prophecy and gender: Can women be Pygmalion and Galatea? *Journal of Applied Psychology*, 80(2), 253–270.
- Eckert, T., Dunn, E. K., Coddling, R. S., Begeny, J. C. & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools*, 43(3), 247–265.
- Eden, D. (1992). Leadership and expectations: Pygmalion effects and other self-fulfilling prophecies in organizations. *The Leadership Quarterly*, 3(4), 271–305. doi:10.1016/1048-9843(92)90018-B
- Ehmke, T. & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231–254). Münster: Waxmann.
- Epstein, J. L. (1988). Effective schools or effective students: Dealing with diversity. In R. Haskins & D. J. MacRae (Hrsg.), *Policies for America's public schools: Teachers, equity, and indicators*. (S. 89–126). Westport, CT, US: Ablex Publishing.
- Feinberg, A. B. & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, 18(1), 52–65. doi:10.1521/scpq.18.1.52.20876
- Feinberg, A. B. & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, 102(6), 453–462.
- Ferreira, M. B., Garcia-Marques, L., Sherman, S. J. & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, 91(5), 797–813. doi:10.1037/0022-3514.91.5.797
- Fiske, S. T., Lin, M. H. & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Hrsg.), *Dual process theories in social psychology* (S. 231–254). New York: Guilford.

- Fiske, S. T. & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Hrsg.), *Advances in experimental social psychology* (Bd. 23, S. 1–74). San Diego: Academic Press Inc.
- Fiske, S. T. & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.
- Fitzpatrick, C., Côté-Lussier, C. & Blair, C. (2016). Dressed and groomed for success in elementary school: Student appearance and academic adjustment. *The Elementary School Journal*, 117(1), 30–45. doi:10.1086/687753
- Förster, N. & Böhmer, I. (2017). Das Linsenmodell – Grundlagen und exemplarische Anwendungen in der pädagogisch-psychologischen Diagnostik. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 46–50). Münster, New York: Waxmann.
- Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K. & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology*, 41, 1–12. doi:10.1016/j.cedpsych.2014.10.006
- Galinsky, A. & Moskowitz, G. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, 78(4), 708–724. doi:10.1037//0022-3514.78.4.708
- Gebhardt, M., Rauch, D., Mang, J., Sälzer, C. & Stanat, P. (2013). Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 275–308). Münster: Waxmann.
- Gentrup, S., Lorenz, G., Kristen, C. & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, 66. doi:10.1016/j.learninstruc.2019.101296
- Gentrup, S. & Rjosk, C. (2018). Pygmalion and the gender gap: Do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling? *Educational Research and Evaluation*, 24(3–5), 295–323. doi:10.1080/13803611.2018.1550840
- Gentrup, S., Rjosk, C., Stanat, P. & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen. *Zeitschrift für Erziehungswissenschaft*, 21, 867–891. doi:10.1007/s11618-018-0806-2
- Gignac, G. E. & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. doi:10.1016/j.paid.2016.06.069

- Gill, S. & Reynolds, A. J. (1999). Educational expectations and school achievement of urban African American children. *Journal of School Psychology, 37*(4), 403–424. doi:10.1016/S0022-4405(99)00027-8
- Givvin, K. B., Stipek, D. J., Salmon, J. M. & MacGyvers, V. L. (2001). In the eyes of the beholder: Students' and teachers' judgments of students' motivation. *Teaching and Teacher Education, 17*(3), 321–331.
- Glock, S. & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Social Psychology of Education, 16*(1), 111–127. doi:10.1007/s11218-012-9197-z
- Glock, S. & Krolak-Schwerdt, S. (2014). Stereotype activation versus application: How teachers process and judge information about students from ethnic minorities and with low socioeconomic background. *Social Psychology of Education, 17*(4), 589–607. doi:10.1007/s11218-014-9266-6
- Good, T. L. (1987). Two decades of research on teacher expectations: Findings and future directions. *Journal of Teacher Education, 38*(4), 32–47. doi:10.1177/002248718703800406
- Good, T. L. & Nichols, S. L. (2001). Expectancy effects in the classroom: A special focus on improving the reading performance of minority students in first-grade classrooms. *Educational Psychologist, 36*(2), 113–126.
- Haag, N., Kocaj, A., Jansen, M. & Kuhl, P. (2017). Soziale Disparitäten. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB-Bildungstrend 2016: Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (S. 213–235). Münster: Waxmann.
- Hachfeld, A., Anders, Y., Schroeder, S., Stanat, P. & Kunter, M. (2010). Does immigration background matter? How teachers' predictions of students' performance relate to student background. *International Journal of Educational Research, 49*(2–3), 78–91.
- Halpern, D. F. (2013). *Sex differences in cognitive abilities*. New York: Psychology Press.
- Hannover, B. (2004). Gender revisited: Konsequenzen aus PISA für die Geschlechterforschung. *Zeitschrift für Erziehungswissenschaft, 7*, 81–99.
- Harris, M. & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin, 97*(3), 363–386. doi:0033-2909/85/S00.75
- Hartley, D. (1982). Ethnicity or sex: Teachers' definitions of ability and reading comprehension in an EPA primary school. *Research in Education, 28*, 9–24.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Hattrup, K. & Ford, J. K. (1995). The roles of information characteristics and accountability in moderating stereotype-driven processes during social decision making. *Organizational Behavior and Human Decision Processes*, 63(1), 73–86. doi:10.1006/obhd.1995.1063
- Hecht, S. A. & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgments of their students' reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing*, 15(7), 789–809.
- Heinze, A., Herwartz-Emden, L. & Reiss, K. (2007). Mathematikkenntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit. *Zeitschrift für Pädagogik*, 53(4), 562–581.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität – Diagnose, Evaluation und Verbesserung des Unterrichts* (7). Seelze-Velber: Klett | Kallmeyer.
- Helmke, A. & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98. doi:10.1016/0742-051X(87)90010-2
- Heyder, A., Steinmayr, R. & Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain gender differences in children's math ability self-concept? *Frontiers in Education*, 4(34). doi:10.3389/educ.2019.00034
- Hill, H. C. & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal*, 55(5), 1076–1112. doi:10.3102/0002831218769614
- Hinnant, J. B., O'Brien, M. & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101(3), 662–670. doi:10.1037/a0014306
- Hoffmann, L. (2017). *Lehrkräfte als Diagnostikerinnen und Diagnostiker. Untersuchungen zu ausgewählten Tätigkeiten von Grundschullehrerinnen und Grundschullehrern im Bereich der pädagogisch-psychologischen Diagnostik*. Dissertation, Universität Potsdam, Potsdam.
- Hoffmann, L. & Böhme, K. (2014). Wie gut können Grundschullehrkräfte die Schwierigkeit von Deutsch- und Mathematikaufgaben beurteilen? Eine Untersuchung zur Genauigkeit aufgabenbezogener Lehrerurteile auf Klassenebene. *Psychologie in Erziehung und Unterricht*, 61, 42–55.
- Hoge, R. D. & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology*, 76(6), 777–781. doi:10.1037//0022-0663.76.5.777

- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313. doi:10.3102/00346543059003297
- Holder, K. & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social Psychology of Education*, 20, 471–490. doi:10.1007/s11218-017-9384-z
- Hollenstein, L., Affolter, B. & Brühwiler, C. (2019). Die Bedeutung der Leistungserwartung von Lehrpersonen für die Mathematikleistungen von Schülerinnen und Schülern. *Zeitschrift für Erziehungswissenschaft*, 22(4), 791–809. doi:10.1007/s11618-019-00901-4
- Hosenfeld, I., Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz. Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (Bd. 45, Suppl., S. 65–82). Weinheim: Beltz. doi:urn:nbn:de:0111-opus-39393
- Hurwitz, J. T., Elliott, S. N. & Braden, J. P. (2007). The influence of test familiarity and student disability status upon teachers' judgments of students' test performance. *School Psychology Quarterly*, 22(2), 115–144.
- Hußmann, A., Stubbe, T. C. & Kasper, D. (2017). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195–217). Münster: Waxmann.
- Impara, J. C. & Plake, B. S. (1998). Teachers' ability to estimate item difficulty: A test of assumptions in the Angoff standard setting method. *Journal of Educational Measurement*, 35(1), 69–81. doi:10.1111/j.1745-3984.1998.tb00528.x
- Jenkins, L. N. & Demaray, M. K. (2016). Teachers' judgments of the academic achievement of children with and without characteristics of inattention, impulsivity, and hyperactivity. *Contemporary School Psychology*, 20(2), 183–191. doi:10.1007/s40688-015-0073-7
- Jungbluth, P. (1993). Pygmalion and the effectiveness of 'Black' schools: Teachers' stereotypes and hidden goal differentiations towards ethnic minorities. *Tijdschrift voor Onderwijsresearch*, 18(2), 99–110.
- Jungbluth, P. (1994). Lehrererwartungen und Ethnizität. Innerschulische Chancendeterminanten bei Migrantenschülern in den Niederlanden. *Zeitschrift für Pädagogik*, 40(1), 113–125. doi:urn:nbn:de:0111-pedocs-108314

- Juska-Bacher, B. (2013). Leserelevante Kompetenzen und ihre frühe Förderung. In M. Stamm & D. Edlmann (Hrsg.), *Handbuch frühkindliche Bildungsforschung* (S. 485–500). Wiesbaden: Springer VS.
- Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review*, 93(4), 429–445. doi:10.1037/0033-295X.93.4.429
- Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology*, 57(3), 469–480. doi:10.1037/0022-3514.57.3.469
- Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological review*, 98(1), 54–73.
- Jussim, L. & Eccles, J. (1992). Teacher expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology*, 63(6), 947–961. doi:10.1037/0022-3514.63.6.947
- Jussim, L. & Eccles, J. (1995). Are teacher expectations biased by students' gender, social class, or ethnicity? In Y.-T. Lee, L. J. Jussim, C. R. McCauley, Y.-T. Lee, L. J. Jussim & C. R. McCauley (Hrsg.), *Stereotype accuracy: Toward appreciating group differences*. (S. 245–271). Washington, DC: American Psychological Association. doi:10.1037/10495-010
- Jussim, L., Eccles, J. & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Hrsg.), *Advances in Experimental Social Psychology* (Bd. 28, S. 281–388). San Diego: Academic Press.
- Jussim, L. & Harber, K. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155. doi:10.1207/s15327957pspr0902_3
- Jussim, L., McCauley, C. R. & Lee, Y.-T. (1995). Why study stereotype accuracy and inaccuracy? In Y.-T. Lee, L. Jussim & C. McCauley (Hrsg.), *Stereotype accuracy: Toward appreciating group differences* (S. 3–27). Washington, DC, US: American Psychological Association. doi:10.1037/10495-001
- Jussim, L., Robustelli, S. & Cain, T. (2009). Teacher expectations and self-fulfilling prophecies. In K. Wentzel & A. Wigfield (Hrsg.), *Handbook of Motivation at School* (S. 349–379). New York: Routledge.
- Kaiser, J., Retelsdorf, J., Südkamp, A. & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction* 28, 73–84.

- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 197–209. doi:10.1024/1010-0652.23.34.197
- Karing, C., Matthäi, J. & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25(3), 159–172. doi:10.1024/1010-0652/a000041
- Keller, E. (2016). *Professionelle Kompetenz angehender Deutsch-Lehrkräfte: Entwicklung und Validierung eines Testverfahrens zur Erfassung fachdidaktischen Wissens im Kompetenzbereich Schreiben*. Dissertation, Pädagogische Hochschule Heidelberg, Heidelberg.
- Kratzmann, J. & Pohlmann-Rother, S. (2012). Ethnische Stereotype im Kindergarten? Erzieherinnenhaltungen gegenüber Zuwanderern aus der Türkei. *Zeitschrift für Pädagogik*, 58(6), 855–876.
- Kristen, C., Stanat, P., Kogan, I., Lorenz, G., Gentrup, S. & Rahmann, S. (2018a). *Kompetenzerwerb und Lernvoraussetzungen (KuL). Test- und Befragungsdaten* [Competence acquisition and learning preconditions (KuL). Assessment and survey data] (Version 1). Data set. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. doi:10.5159/IQB_KuL_v1
- Kristen, C., Stanat, P., Kogan, I., Lorenz, G., Gentrup, S. & Rahmann, S. (2018b). *Kompetenzerwerb und Lernvoraussetzungen (KuL). Videodaten* [Competence acquisition and learning preconditions (KuL). Observational data] (Version 1). Data set. Forschungsdatenzentrum Bildung am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. doi:10.7477/287:1:0
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. Der Lehrer als "flexibler Denker". *Zeitschrift für pädagogische Psychologie*, 23(3–4), 175–186.
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2012). Leistungsbeurteilungen von Schulkindern. Welche Rolle spielen Ziele und Expertise der Lehrkraft? *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 44(3), 111–122. doi:10.1026/0049-8637/a000062
- Krolak-Schwerdt, S. & Rummer, R. (2005). Der Einfluss von Expertise auf den Prozess der schulischen Leistungsbeurteilung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37(4), 205–213. doi:10.1026/0049-8637.37.4.205
- Kuhl, P. & Hannover, B. (2012). Differenzielle Benotungen von Mädchen und Jungen. Der Einfluss der von der Lehrkraft eingeschätzten Kompetenz zum selbstgesteuerten Lernen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(3), 153–162. doi:10.1026/0049-8637/a000066

- Kuklinski, M. R. & Weinstein, R. S. (2000). Classroom and grade level differences in the stability of teacher expectations and perceived differential teacher treatment. *Learning Environments Research*, 3, 1–34.
- Kuklinski, M. R. & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development*, 72(5), 1554–1578.
- Kunina-Habenicht, O., Decker, A. & Kunter, M. (2015). Lehrerpersönlichkeit und professionelle Kompetenzen von Lehrkräften. In K. Seifried, S. Drewes & M. Hasselhorn (Hrsg.), *Handbuch Schulpsychologie: Psychologie für die Schule* (S. 319–330). Stuttgart: Kohlhammer.
- Kunter, M. & Pohlmann, B. (2015). Lehrer. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 261–281). Berlin: Springer. doi:10.1007/978-3-642-41291-2_11
- Li, Z. & Rubie-Davies, C. M. (2017). Teachers matter: Expectancy effects in Chinese university English-as-a-foreign-language classrooms. *Studies in Higher Education*, 42(11), 2042–2060. doi:10.1080/03075079.2015.1130692
- Liu, W. C. & Wang, C. K. J. (2008). Home environment and classroom climate: An investigation of their relation to students' academic self-concept in a streamed setting. *Current Psychology*, 27(4), 242. doi:10.1007/s12144-008-9037-7
- Lorenz, C. (2011). *Diagnostische Kompetenz von Grundschullehrkräften – Strukturelle Aspekte und Bedingungen* (Bd. 9). Bamberg: University of Bamberg Press.
- Lorenz, C. & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 211–222. doi:10.1024/1010-0652.23.34.211
- Lorenz, G. (2018). *Selbsterfüllende Prophezeiungen in der Schule. Leistungserwartungen von Lehrkräften und Kompetenzen von Kindern mit Zuwanderungshintergrund*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lorenz, G. (2019). Subtle discrimination? Stereotypes among teachers trigger ethnic bias in their achievement expectations but are unrelated to ethnic achievement gaps. Manuskript in Vorbereitung.
- Lorenz, G., Gentrup, S., Kristen, C., Stanat, P. & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(1), 89–111. doi:10.1007/s11577-015-0352-3
- Ludwig, P. H. (2007). Pygmalion zwischen Venus und Mars – Geschlechtsunterschiede in schulischen Lernleistungen durch Selbsterfüllung von Erwartungen. In P. H. Ludwig & H. Ludwig (Hrsg.), *Erwartungen in himmelblau und rosarot* (S. 17–59). Weinheim: Juventa.

- Lüftenegger, M., van de Schoot, R., Schober, B., Finsterwald, M. & Spiel, C. (2014). Promotion of students' mastery goal orientations: Does TARGET work? *Educational Psychology*, 34(4), 451–469. doi:10.1080/01443410.2013.814189
- Madelaine, A. & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education*, 52(1), 33–42.
- Madon, S., Jussim, L. & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of personality and social psychology*, 72(4), 791–809. doi:10.1037/0022-3514.72.4.791
- Madon, S., Jussim, L. J., Keiper, S., Eccles, J., Smith, A. E. & Palumbo, P. (1998). The accuracy and power of sex, social class, and ethnic stereotypes: A naturalistic study in person perception. *Personality and Social Psychology Bulletin*, 24(12), 1304–1318.
- Matsuoka, R. (2014). Disparities between schools in Japanese compulsory education: Analyses of a cohort using TIMSS 2007 and 2011. *Educational Studies in Japan*, 8, 77–92. doi:10.7571/esjkyoiku.8.77
- McElvany, N., Kessels, U., Schwabe, F. & Kasper, D. (2017). Geschlecht und Lesekompetenz. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 177–194). Münster: Waxmann.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W. et al. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23(34), 223–235. doi:10.1024/1010-0652.23.34.223
- McKown, C. & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology*, 32(1), 159–184. doi:10.1111/j.1559-1816.2002.tb01425.x
- McKown, C. & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, 46(3), 235–261. doi:10.1016/j.jsp.2007.05.001
- McNatt, D. B. (2000). Ancient Pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, 85(2), 314–322. doi:10.1037/0021-9010.85.2.314
- Meissel, K., Meyer, F., Yao, E. S. & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. doi:10.1016/j.tate.2017.02.021
- Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, 8(2), 193–210. doi:10.2307/4609267
- Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelsen Scriptor.

- Montague, M. & Rinaldi, C. (2001). Classroom dynamics and children at risk: A followup. *Learning Disability Quarterly*, 24(2), 75–83. doi:10.2307/1511063
- Müller, K. & Ehmke, T. (2013). Soziale Herkunft als Bedingung der Kompetenzentwicklung. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 245–274). Münster: Waxmann.
- Muntoni, F. & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading – The role of teachers' gender stereotypes. *Contemporary Educational Psychology*, 54, 212–220. doi:10.1016/j.cedpsych.2018.06.012
- Naumann, J., Artelt, C., Schneider, W. & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23–71). Münster: Waxmann.
- Nestler, S. & Back, M. D. (2013). Applications and extensions of the lens model to understand interpersonal judgments at zero acquaintance. *Current Directions in Psychological Science*, 22(5), 374–379. doi:10.1177/0963721413486148
- Nölle, I., Hörstermann, T., Krolak-Schwerdt, S. & Gräsel, C. (2009). Relevante diagnostische Informationen bei der Übergangsempfehlung – Die Perspektive der Lehrkräfte. *Unterrichtswissenschaft*, 37(4), 294–310.
- Oerke, B., McElvany, N., Ohle, A., Ullrich, M. & Horz, H. (2016). Verbessert sich die diagnostische Urteilsgenauigkeit von Lehrkräften bei längerem Kontakt mit der Klasse? *Psychologie in Erziehung und Unterricht*, 63(1), 34–47.
- Ophuysen, S. v. (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahneempfehlung. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 38(4), 154–161.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332. doi:10.3102/00346543062003307
- Pendry, L. (2014). Soziale Kognition. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Sozialpsychologie* (Bd. 6, S. 107–140). Heidelberg: Springer.
- Pendry, L. & Macrae, C. N. (1996). What the disinterested perceiver overlooks: Goal-directed social categorization. 22(3), 249–256. doi:10.1177/0146167296223003
- Pesu, L., Viljaranta, J. & Aunola, K. (2016). The role of parents' and teachers' beliefs in children's self-concept development. *Journal of Applied Developmental Psychology*, 44, 63–71. doi:10.1016/j.appdev.2016.03.001
- Peterson, E. R., Rubie-Davies, C. M., Osborne, D. & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, 42, 123–140.

- Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A. & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging self-concepts of students. *The Journal of Educational Research*, 106(1), 64–76.
- Praetorius, A.-K., Drexler, K., Rösch, L., Christophel, E., Heyne, N., Scheunpflug, A. et al. (2015). Judging students' self-concepts within 30 s? Investigating judgement accuracy in a zero-acquaintance situation. *Learning and Individual Differences*, 37, 231–236. doi:10.1016/j.lindif.2014.11.015
- Praetorius, A.-K. & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A.-K. Praetorius & A. Südkamp (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 13–18). Münster, New York: Waxmann.
- Praetorius, A.-K. & Südkamp, A. (2019). Diagnostische Kompetenz von Lehrpersonen hinsichtlich motivationaler Merkmale von Schülerinnen und Schülern. In H. Gaspard, U. Trautwein & M. Hasselhorn (Hrsg.), *Diagnostik und Förderung von Motivation und Volition* (1. Auflage ed.). Göttingen: Hogrefe.
- Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (Hrsg.). (2013). *PISA 2012. Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.
- Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, 76(1), 85–97. doi:10.1037/0022-0663.76.1.85
- Ready, D. D. & Chu, E. M. (2015). Sociodemographic inequality in early literacy development: The role of teacher perceptual accuracy. *Early Education and Development*, 26(7), 970–987. doi:10.1080/10409289.2015.1004516
- Ready, D. D. & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, 48(2), 335–360. doi:10.3102/0002831210374874
- Rellensmann, J. & Schukajlow, S. (2017). Does students' interest in a mathematical problem depend on the problem's connection to reality? An analysis of students' interest and pre-service teachers' judgments of students' interest in problems with and without a connection to reality. 49(3), 367–378. doi:10.1007/s11858-016-0819-3
- Riegle-Crumb, C. & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender and Society*, 26(2), 290–322. doi:10.1177/0891243211434614

- Rjosk, C., Haag, N., Heppt, B. & Stanat, P. (2017). Zuwanderungsbezogene Disparitäten. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB-Bildungstrend 2016: Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (S. 237–275). Münster: Waxmann.
- Rjosk, C., McElvany, N., Anders, Y. & Becker, M. (2011). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung der basalen Lesefähigkeit ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 58(2), 92–105.
- Robinson-Cimpian, J. P., Theule Lubienski, S., Ganley, C. M. & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology*, 50(4), 1262–1281. doi:10.1037/a0035073
- Rosenthal, R. (1974). *On the social psychology of the self-fulfilling prophecy: Further evidence for pygmalion effects and their mediating mechanisms* (Bd. Module 53). New York: MSS Modular Publications.
- Rosenthal, R. (1989). Experimenter expectancy, covert communication, and meta-analytic methods, *Paper presented at the Annual Meeting of the American Psychological Association*. New Orleans.
- Rosenthal, R. & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3, 16–20.
- Rosenthal, R. & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 1(3), 377–386. doi:10.1017/S0140525X00075506
- Ross, S. I. & Jackson, J. M. (1991). Teachers' expectations for Black males' and Black females' academic achievement. *Personality and Social Psychology Bulletin*, 17(1), 78–82. doi:10.1177/0146167291171012
- Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology*, 77(2), 289–306. doi:10.1348/000709906X101601
- Rubie-Davies, C. M., Hattie, J. A. C., Townsend, M. A. R. & Hamilton, R. J. (2007). Aiming high: Teachers and their students. In V. N. Galwey (Hrsg.), *Progress in Educational Psychology Research* (S. 65–91). Hauppauge, NY: Nova Publisher.
- Rubie-Davies, C. M. & Peterson, E. R. (2016). Relations between teachers' achievement, over- and underestimation, and students' beliefs for Māori and Pākehā students. *Contemporary Educational Psychology*, 47, 72–83. doi:10.1016/j.cedpsych.2016.01.001
- Rubie-Davies, C. M., Peterson, E. R., Sibley, C. G. & Rosenthal, R. (2015). A teacher expectation intervention: Modelling the practices of high expectation teachers. *Contemporary Educational Psychology*, 40, 72–85. doi:10.1016/j.cedpsych.2014.03.003

- Rubie-Davies, C. M., Weinstein, R. S., Huang, F. L., Gregory, A., Cowan, P. A. & Cowan, C. P. (2014). Successive teacher expectation effects across the early school years. *Journal of Applied Developmental Psychology*, 35(3), 181–191. doi:10.1016/j.appdev.2014.03.006
- Rubie, C. M. (2003). *Expecting the best: Instructional practices, teacher beliefs and student outcomes*. Dissertation, The University of Auckland, Auckland.
- Rudman, L., Ashmore, R. & Gary, M. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of personality and social psychology*, 81(5), 856–868. doi:10.1037//0022-3514.81.5.856
- Sälzer, C., Reiss, K., Schiepe-Tiska, A., Prenzel, M. & Heinze, A. (2013). Zwischen Grundlagenwissen und Anwendungsbezug: Mathematische Kompetenz im internationalen Vergleich. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 47–98). Münster: Waxmann.
- Schenke, K., Nguyen, T., Watts, T. W., Sarama, J. & Clements, D. H. (2017). Differential effects of the classroom on African American and non-African American's mathematics achievement. *Journal of Educational Psychology*. doi:10.1037/edu0000165
- Schipolowski, S., Wittig, J., Weirich, S. & Böhme, K. (2017). Geschlechtsbezogene Disparitäten. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (S. 187–211). Münster: Waxmann.
- Schneider, T. (2011). Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschulempfehlung. *Zeitschrift für Erziehungswissenschaft*, 14(3), 371–396. doi:10.1007/s11618-011-0221-4
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt, Main: Lang.
- Schrader, F.-W. (2010). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 102–108). Weinheim: Beltz Verlag.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik: Zeitschrift zu Theorie und Praxis erziehungswissenschaftlicher Forschung*, 1(1), 27–42.
- Schrader, F.-W. & Helmke, A. (1990). Lassen sich Lehrer bei der Leistungsbeurteilung von sachfremden Gesichtspunkten leiten? Eine Untersuchung zu Determinanten diagnostischer Lehrerurteile. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22(4), 312–324.
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Weinheim: Beltz.

- Segeritz, M., Walter, O. & Stanat, P. (2010). Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland: Evidenz für segmentierte Assimilation? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62, 113–138.
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik*, 60(6), 850–866.
- Seidel, T. & Reiss, K. (2014). Lerngelegenheiten im Unterricht. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie* (S. 253–276). Weinheim: Beltz.
- Shifrer, D. (2013). Stigma of a label: Educational expectations for high school students labeled with learning disabilities. *Journal of Health and Social Behavior*, 54(4), 462–480. doi:10.1177/0022146513503346
- Shifrer, D. (2016). Stigma and stratification limiting the math course progression of adolescents labeled with a learning disability. *Learning and Instruction*, 42, 47–57. doi:10.1016/j.learninstruc.2015.12.001
- Snow, R. E. (1969). Review of "Pygmalion in the classroom" by Rosenthal and Jacobsen. *Current Directions in Psychological Science*, 14, 197–199.
- Snow, R. E. (1995). Pygmalion and intelligence? *Current Directions in Psychological Science*, 4(6), 169–171. doi:10.1111/1467-8721.ep10772605
- Sorhagen, N. S. (2013). Early teacher expectations disproportionately affect poor children's high school performance. *Journal of Educational Psychology*, 105(2), 465–477. doi:10.1037/a0031754
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 85–95. doi:10.1024/1010-0652.19.1.85
- Spinath, B. (2011). Lernmotivation. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche* (S. 45–55). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stahl, N. (2007). Schülerwahrnehmung und -beurteilung durch Lehrkräfte. In H. Ditton (Hrsg.), *Kompetenzaufbau und Laufbahnen im Schulsystem* (S. 171–198). Münster: Waxmann.
- Stanat, P., Bergann, S. & Taraszow, T. (2018). Geschlechtsbezogene Disparitäten im deutschen Bildungswesen. In R. Tippelt & B. Schmidt-Hertha (Hrsg.), *Handbuch Bildungsforschung* (S. 1321–1338). Wiesbaden: Springer Fachmedien. doi:10.1007/978-3-531-19981-8_59
- Stanat, P., Rauch, D. & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200–230). Münster: Waxmann. doi:urn:nbn:de:0111-opus-35360

- Stubbe, T. C., Schwippert, K. & Wendt, H. (2016). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 299–316). Münster: Waxmann.
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. doi:10.1037/a0027627
- Südkamp, A. & Möller, J. (2009). Referenzgruppeneffekte im Simulierten Klassenraum: Direkte und indirekte Einschätzungen von Schülerleistungen. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 161–174.
- Südkamp, A., Möller, J. & Pohlmann, B. (2008). Der simulierte Klassenraum: Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22(3–4), 261–276.
- Tenenbaum, H. R. & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology*, 99(2), 253–273. doi:10.1037/0022-0663.99.2.253
- Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. P. Zanna (Hrsg.), *Advances in Experimental Social Psychology* (Bd. 25, S. 331–376): Academic Press. doi:10.1016/S0065-2601(08)60287-7
- Tetlock, P. E. & Lerner, J. S. (1999). The social contingency model: Identifying empirical and normative boundary conditions on the error-and-bias portrait of human nature. In S. Chaiken & Y. Trope (Hrsg.), *Dual-process theories in social psychology* (S. 571–585). New York, NY, US: Guilford Press.
- Thorndike, R. L. (1968). Review of Robert Rosenthal and Lenore Jacobson, *Pygmalion in the classroom*. *American Educational Research Journal*, 5(4), 708–711. doi:10.2307/1162010
- Tiedemann, J. (2000). Gender-related beliefs of teachers in elementary school mathematics. *Educational Studies in Mathematics*, 41(2), 191–207.
- Tiedemann, J. (2002). Teacher's gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics*, 50, 49–62.
- Timmermans, A. C., de Boer, H. & van der Werf, M. P. C. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education*, 19(2), 217–240. doi:10.1007/s11218-015-9326-6

- Timmermans, A. C., Kuyper, H. & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, 85(4), 459–478. doi:10.1111/bjep.12087
- Tobisch, A. & Dresel, M. (2017). Negatively or positively biased? Dependencies of teachers' judgments and expectations based on students' ethnic and social backgrounds. *Social Psychology of Education*, 20(4), 731–752. doi:10.1007/s11218-017-9392-z
- Trautwein, U. & Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind: Referenzgruppeneffekte bei Übertrittsentscheidungen. *Zeitschrift für Pädagogische Psychologie*, 21(2), 119–133.
- Trouilloud, D. & Sarrazin, P. (2003). Les connaissances actuelles sur l'effet Pygmalion: Processus, poids et modulateurs. *Revue Française de Pédagogie*, 145, 89–119. doi:10.3406/rfp.2003.2988
- Trouilloud, D., Sarrazin, P., Martinek, T. & Guillet, E. (2002). The influence of teacher expectations on student achievement in physical education classes: Pygmalion revisited. *European Journal of Social Psychology*, 32(5), 591–607. doi:10.1002/ejsp.109
- Trusz, S. (2018). Four mediation models of teacher expectancy effects on students' outcomes in mathematics and literacy. *Social Psychology of Education*, 21(2), 257–287. doi:10.1007/s11218-017-9418-6
- Tschannen-Moran, M. & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783–805. doi:10.1016/S0742-051X(01)00036-1
- Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education*, 45, 73–82. doi:10.1016/j.tate.2014.09.006
- Urhahne, D., Chao, S.-H., Florineth, M. L., Luttenberger, S. & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student. *British Journal of Educational Psychology*, 81, 161–177.
- Urhahne, D., Zhou, J., Stobbe, M., Chao, S.-H., Zhu, M. & Shi, J. (2010). Motivationale und affektive Merkmale unterschätzter Schüler. *Zeitschrift für Pädagogische Psychologie*, 24(3–4), 275–288.
- Urhahne, D. & Zhu, M. (2015). Accuracy of teachers' judgments of students' subjective well-being. *Learning & Individual Differences*, 43, 226–232. doi:10.1016/j.lindif.2015.08.007
- van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M. & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, 47(2), 497–527. doi:10.3102/0002831209353594

- Van Houtte, M., Demanet, J. & Stevens, P. A. J. (2013). Curriculum tracking and teacher evaluations of individual students: Selection, adjustment or labeling? *Social Psychology of Education*, 16(3), 329–352. doi:10.1007/s11218-013-9216-8
- van Matre, J. C., Valentine, J. C. & Cooper, H. M. (2000). Effect of students' after-school activities on teachers' academic expectancies. *Contemporary Educational Psychology*, 25, 167–183. doi:10.1006/ceps.1998.0999
- Walter, O. & Taskinen, P. (2008). Naturwissenschaftsbezogene Motivationen und Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund in Deutschland. Der Einfluss der Generationen, der Herkunft und des Elternhauses. In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006* (S. 185–203). Wiesbaden: Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-91815-0_10
- Wang, S., Rubie-Davies, C. M. & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, 24(3–5), 124–179. doi:10.1080/13803611.2018.1548798
- Wanzek, J., Roberts, G. & Al Otaiba, S. (2014). Academic responding during instruction and reading outcomes for kindergarten students at-risk for reading difficulties. *Reading and Writing*, 27(1), 55–78. doi:10.1007/s11145-013-9433-8
- Watson, P., Rubie-Davies, C. M., Meissel, K., Peterson, E., Flint, A., Garrett, L. et al. (2014). Gendered teacher expectations of mathematics achievement in New Zealand: Contributing to a kink at the base of the STEM pipeline. *International Journal of Gender, Science and Technology*, 8(1), 83–102.
- Weinert, F. E. & Schrader, F.-W. (1986). Diagnose des Lehrers als Diagnostiker. In H. Petillon, J. W. L. Wagner & B. Wolf (Hrsg.), *Schülergerechte Diagnose: Theoretische und empirische Beiträge zur Pädagogischen Diagnostik* (S. 11–29). Weinheim: Beltz.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990). Educational expertise: Closing the gap between educational research and classroom practice. *School Psychology International*, 11(3), 163–180.
- Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling*: Harvard University Press.
- Wendt, H., Steinmayr, R. & Kasper, D. (2016). Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 257–298). Münster: Waxmann.

- Westphal, A., Gronostaj, A., Vock, M., Emmrich, R. & Harych, P. (2016). Differenzierung im gymnasialen Mathematik und Deutschunterricht – vor allem bei guten Diagnostiker/innen und in heterogenen Klassen? *Zeitschrift für Pädagogik*, 62, 131–148.
- Wild, K.-P. & Rost, D. H. (1995). Klassengröße und Genauigkeit von Schülerbeurteilungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 27(1), 78–90.
- Wood, D., Kaplan, R. & McLoyd, V. C. (2007). Gender differences in the educational expectations of urban, low-income african american youth: The role of parents and the school. *Journal of Youth and Adolescence*, 36(4), 417–427. doi:10.1007/s10964-007-9186-2
- Woolfolk Hoy, A., Davis, H. & Pape, S. J. (2006). Teacher knowledge and beliefs. In P. A. Alexander & P. H. Winne (Hrsg.), *Handbook of educational psychology* (S. 715–737). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Zhou, J. & Urhahne, D. (2013). Teacher judgment, student motivation, and the mediating effect of attributions. *European Journal of Psychology of Education*, 28(2), 275–295.
- Zhu, M., Urhahne, D. & Rubie-Davies, C. M. (2018). The longitudinal effects of teacher judgement and different teacher treatment on students' academic outcomes. *Educational Psychology*, 38(5), 648–668. doi:10.1080/01443410.2017.1412399

Schlussteil der Arbeit

Abstract

The present dissertation deals with teacher expectations for student achievement. The first two studies investigate student characteristics teachers rely on when forming their achievement expectations. Study 1 analyses whether teachers' expectations of students' achievement are biased by students' ethnic background, social background and gender. Study 2 asks whether such biases in teachers' achievement expectations are partly explained by differences in teachers' perceptions of students' motivation and learning behavior. Study 3 investigates the associations of inaccurately high or low teacher expectations with their teaching behavior and students' subsequent achievement development. The analyses for these research questions are based on data from the longitudinal research project "competence acquisition and learning preconditions" in which 1065 students from 64 first grade classes and their teachers participated in the school year 2013/2014.

The multilevel analyses of study 1 revealed that teachers' expectations for students' language and mathematics achievement differed depending on students' ethnic background, social background and gender. These differences partly persisted even after controlling for students' actual achievement, general cognitive abilities and motivation, and therefore indicate biased expectations. The results of study 2 showed that social bias and gender bias in teacher expectations were partly due to differences in teachers' perceptions of students' motivation and learning behavior. Ethnic bias in teacher expectations, however, appeared independent of teachers' perceptions of students' motivation and learning behavior. The results of study 3 support the assumption that inaccurately high or low teacher expectations may result in self-fulfilling prophecies and influence students' achievement development as early as in the first school grade. The video study of study 3 further revealed that inaccurately high or low teacher expectations may result in different teacher feedback for the students. Teacher feedback, however, did not substantially mediate teacher expectancy effects on student achievement.

The results of the present dissertation are of great relevance for educational practice. First, they highlight the advantages of high expectations and, second, they sensitize for biasing influences of student background characteristics.

Zusammenfassung

Die vorliegende Dissertation untersucht in drei Teilstudien Leistungserwartungen von Lehrkräften. Die ersten beiden Teilstudien befassen sich mit der Frage, welche Merkmale von Schüler*innen Lehrkräfte in ihre Erwartungen an die Leistungen von Lernenden einbeziehen. Teilstudie 1 untersucht dabei, ob Leistungserwartungen von Lehrkräften in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund und Geschlecht der Schüler*innen verzerrt sind. Daran anknüpfend fragt Teilstudie 2, ob solche Verzerrungen zum Teil auf Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens der Schüler*innen zurückgeführt werden können. Welche Bedeutung inakkurat hohe oder niedrige Leistungserwartungen von Lehrkräften für das unterrichtliche Handeln der Lehrkräfte und letztlich für die Leistungsentwicklungen der Schüler*innen haben, wird in Teilstudie 3 untersucht. Zur Analyse dieser Fragestellungen dienen Daten der Längsschnittstudie „Kompetenzerwerb und Lernvoraussetzungen“ (KuL), an der im Schuljahr 2013/2014 insgesamt 1065 Schüler*innen aus 64 ersten Grundschulklassen und ihre Lehrkräfte teilgenommen haben.

Die Mehrebenenanalysen in Teilstudie 1 ergaben, dass die Erwartungen von Lehrkräften an die sprachlichen und mathematischen Leistungen der Erstklässler*innen in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund und Geschlecht der Lernenden variieren. Diese Unterschiede blieben zum Teil auch nach Kontrolle der gemessenen Leistungen, allgemeinen kognitiven Fähigkeiten und selbsteingeschätzten Motivation der Schüler*innen bestehen und kennzeichnen daher Verzerrungen. Den Ergebnissen von Teilstudie 2 zufolge gehen solche sozialen und geschlechtsbezogenen Verzerrungen in Leistungserwartungen zum Teil auf die Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens zurück. Ethnische Verzerrungen in Leistungserwartungen scheinen hingegen unabhängig von den Einschätzungen der Motivation und des Arbeitsverhaltens zu bestehen. Dass inakkurat hohe oder niedrige Leistungserwartungen die Leistungsentwicklung von Lernenden bereits im ersten Grundschuljahr im Sinne einer selbsterfüllenden Prophezeiung beeinflussen können, zeigten die Ergebnisse der Teilstudie 3. Die zugehörige Videostudie verdeutlichte zudem, dass sich inakkurat hohe bzw. niedrige Leistungserwartungen der Lehrkräfte in unterschiedlichem Feedback für die Lernenden niederschlagen. Eine bedeutsame Mediation der Erwartungseffekte auf die Leistungsentwicklung der Schüler*innen durch das Feedback ließ sich allerdings nicht nachweisen.

Die Befunde der vorliegenden Dissertation sind von hoher Relevanz für die pädagogische Praxis. Sie sprechen einerseits für den Vorteil hoher Erwartungen und sensibilisieren andererseits für das Risiko verzerrender Einflüsse von Hintergrundmerkmalen der Schüler*innen.

Teilstudie 1: Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen

Dieser Beitrag ist in der *Kölner Zeitschrift für Soziologie und Sozialpsychologie* erschienen. Die Referenz lautet:

Lorenz, G., Gentrup, S., Kristen, C., Stanat, P. & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen [Stereotypes among teachers? A study of systematic bias in teacher expectations]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(1), 89–111. Doi: 10.1007/s11577-015-0352-3

The final publication is available at Springer Nature via

<http://dx.doi.org/10.1007/s11577-015-0352-3>

Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen

**Georg Lorenz · Sarah Gentrup · Cornelia Kristen ·
Petra Stanat · Irena Kogan**

© Springer Fachmedien Wiesbaden 2016

Zusammenfassung: Die vorliegende Studie geht der Frage nach, ob Lehrererwartungen unabhängig von den Kompetenzen der Schüler nach dem Geschlecht, nach der sozialen oder nach der ethnischen Herkunft variieren. Systematische Verzerrungen liegen dann vor, wenn Lehrkräfte die zukünftigen Leistungen in Abhängigkeit askriptiver Merkmale unter- oder überschätzen. Derartige Muster könnten durch unzutreffende Stereotype zustande kommen. Vor diesem Hintergrund werden die Leistungserwartungen von 69 Grundschullehrkräften untersucht. Die Ergebnisse belegen, dass die zu Schuljahresbeginn geäußerten Leistungserwartungen im Fach Deutsch bei gleichen Ergebnissen in objektiven Leistungstests für Kinder mit einem türkischen Zuwanderungshintergrund, für Kinder aus sozial schwächeren

Ein Hinweis vorab: Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für beiderlei Geschlecht.

G. Lorenz (✉) · C. Kristen
Lehrstuhl für Soziologie, Otto-Friedrich-Universität Bamberg,
Feldkirchenstr. 21, 96045 Bamberg, Deutschland
E-Mail: georg.lorenz@uni-bamberg.de

C. Kristen
E-Mail: cornelia.kristen@uni-bamberg.de

S. Gentrup · P. Stanat
Humboldt-Universität Berlin / IQB,
Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: sarah.gentrup@iqb.hu-berlin.de

P. Stanat
E-Mail: petra.stanat@iqb.hu-berlin.de

I. Kogan
Universität Mannheim / MZES,
68131 Mannheim, Deutschland
E-Mail: irena.kogan@uni-mannheim.de

Familien sowie für Jungen negativ verzerrt sind. Bei der Betrachtung des Unterrichtsfachs Mathematik finden sich positive Verzerrungen für Kinder mit einem osteuropäischen Zuwanderungshintergrund und für Kinder aus sozial besser gestellten Familien. Geschlechtsspezifische Fehleinschätzungen im Fachbereich Mathematik sind nicht erkennbar.

Schlüsselwörter: Lehrererwartungen · Geschlecht · Ethnische Herkunft · Soziale Herkunft · Bildungsungleichheit · Stereotype

Stereotypes Among Teachers? A Study of Systematic Bias in Teacher Expectations

Abstract: Teacher expectations towards certain groups may be systematically biased, as ascriptive characteristics can trigger group-specific stereotypes. As a result teachers may under- or overestimate students' future achievements based on ascriptive characteristics. This study investigates whether teacher expectations are systematically biased towards students' gender, their social or ethnic origin. The empirical analyses are based on a study of 69 primary school teachers who were asked to assess the future performance of first graders in the school subjects German and Mathematics. The findings show that immediately after the start of the school year teacher expectations in the school subject German are negatively biased when it comes to students of Turkish origin, to children of lower social origin and to boys. In the mathematical field, teachers erroneously predict higher achievements for children of Eastern European origin and for those stemming from socioeconomically advantaged families. Gender bias in Mathematics, however, does not seem to occur.

Keywords: Teacher expectations · Gender · Ethnic origin · Social origin · Educational inequalities · Stereotypes

1 Einleitung

Ungleiche Chancen der schulischen Kompetenzentwicklung – nach sozialer oder ethnischer Herkunft und Geschlecht – zählen nach wie vor zu den zentralen Themen der Bildungsforschung. Diese Bildungsungleichheiten lassen sich im Falle sozialer und ethnischer Disparitäten vor allem über Unterschiede in der Ressourcenausstattung der Lernumwelten innerhalb der Familie, aber auch in der Schule und im sozialen Umfeld erklären (Baumert et al. 2006; Esser 2006; Kristen und Granato 2007; Kristen 2008; Müller und Stanat 2006; Stanat und Christensen 2006). Geschlechterunterschiede werden häufig mit geschlechtsspezifischen Sozialisationsprozessen in Verbindung gebracht (Halpern 2013; Hannover 2004; Stanat und Bergann 2009).

Während diese Aspekte in derzeitigen Analysen regelmäßig im Vordergrund stehen, werden Diskriminierungsprozesse im schulischen Kontext nur vereinzelt

behandelt (vgl. Diehl und Fick 2016). Bisherige Analysen zeigen, dass Noten und Übergangsempfehlungen bezogen auf die soziale Herkunft von Schülern teilweise unzutreffend sind (Dumont et al. 2014; Stubbe und Bos 2008). In Bezug auf mögliche Diskriminierungen gegenüber Kindern mit Zuwanderungshintergrund im Rahmen der Notenvergabe und der Übergangsempfehlungen liegen widersprüchliche Befunde vor (Arnold et al. 2007; Bos et al. 2004; Gresch 2012; Lüdemann und Schwerdt 2013; Sprietsma 2013). Allerdings können die betreffenden Studien nicht prüfen, ob bestehende Unterschiede in den vorangehenden schulischen Leistungen mit vorgelagerten Diskriminierungen in Verbindung stehen.

In diesem Zusammenhang kommen Einflüsse der Erwartungen von Lehrkräften an die zukünftigen Leistungen ihrer Schüler in Betracht. So legen vor allem aus den USA stammende Arbeiten nahe, dass Lehrererwartungen Auswirkungen auf die Kompetenzentwicklung haben (vgl. Jussim et al. 2009; Rosenthal und Jacobson 1968). Allerdings wurden Erwartungseffekte bislang nur vereinzelt explizit mit der Entstehung von Bildungsungleichheiten in Verbindung gebracht (vgl. Ready und Chu 2015). Dies setzt eine fundierte Untersuchung der Lehrererwartungen selbst voraus. So müsste zunächst gezeigt werden, dass die Leistungserwartungen der Lehrkräfte für Kinder mit einer unterschiedlichen sozialen Herkunft oder einer unterschiedlichen ethnischen Herkunft oder für Jungen und Mädchen variieren. Darüber hinaus müssten diese wahrgenommenen Differenzen im Verhältnis zu den tatsächlichen Gruppenunterschieden in den schulischen Leistungen zu hoch oder zu niedrig und damit systematisch verzerrt sein. Erwartungen, die gegenüber einzelnen Schülern irrtümlicherweise zu hoch oder zu niedrig ausfallen, können zwar zu individuellen Vor- oder Nachteilen führen, nicht jedoch zu einer Verstärkung von Ungleichheiten auf Gruppenebene.

Einige Studien aus dem internationalen Kontext deuten darauf hin, dass Lehrkräfte Kindern aus sozial schwächeren Familien und aus Zuwandererfamilien geringere Erwartungen entgegenbringen (Jussim und Harber 2005). Auch an Jungen und Mädchen werden je nach beobachtetem Kompetenzbereich jeweils höhere oder niedrigere Erwartungen gestellt (Hinnant et al. 2009; McKown und Weinstein 2002). Allerdings ist die Befundlage zu der Frage, ob Unterschiede in den Einschätzungen gemessen an den schulischen Leistungen der Kinder eher zutreffend oder eher verzerrt sind, heterogen (Jussim et al. 2009). So finden sich sowohl für zutreffende Erwartungen als auch für systematische Verzerrungen Belege. Eine Ursache für die gemischten Befunde könnte darin liegen, dass die den Studien zugrunde liegenden Daten oftmals keine geeigneten Maße für Lehrererwartungen enthalten. Dies betrifft in erster Linie die zur Verfügung stehenden Indikatoren für Lehrererwartungen sowie den Zeitpunkt der Messung dieser Indikatoren im Verlauf der Bildungskarrieren der Kinder.

Für den deutschen Kontext liegen keine ausreichenden empirischen Befunde zum Zusammenhang zwischen Lehrererwartungen und askriptiven Schülermerkmalen vor. Auch Stereotype oder Vorurteile von Lehrkräften, welche als primäre Ursache von systematischen Unterschieden in den Erwartungen an verschiedene Schülergruppen betrachtet werden können (Jussim et al. 1996), wurden in Deutschland bislang kaum untersucht (Hirschauer und Kullmann 2010, S. 353 ff.).

Der vorliegende Beitrag soll deshalb klären, ob die Leistungserwartungen von Grundschullehrkräften in Deutschland über tatsächliche Kompetenzen hinaus mit

der ethnischen Herkunft, der sozialen Herkunft oder dem Geschlecht der Lernenden assoziiert sind. Dabei sollen einerseits die messmethodischen Defizite internationaler Studien überwunden werden. Andererseits soll zusätzlich geklärt werden, ob fachbereichsspezifische Unterschiede in Bezug auf das Ausmaß an systematischen Verzerrungen bestehen und ob Verzerrungen nach der sozialen und der ethnischen Herkunft sowie nach dem Geschlecht der Kinder unabhängig voneinander auftreten. In Abschn. 2 stehen zunächst theoretische Überlegungen zum Entstehungsprozess systematisch verzerrter Lehrererwartungen im Vordergrund. Abschnitt 3 nimmt den Forschungsstand sowie die typischerweise genutzten Messmodelle in den Blick. In Abschn. 4 und 5 werden die Datengrundlage und die Ergebnisse der empirischen Analysen beschrieben. Die Befunde werden in Abschn. 6 zusammengefasst und diskutiert.

2 Theoretische Annahmen über die Entstehung systematisch verzerrter Lehrererwartungen

Anhaltspunkte für eine Erklärung systematisch verzerrter Lehrererwartungen liefert der Ansatz der statistischen Diskriminierung (Phelps 1972; Arrow 1973). Dieser geht davon aus, dass Akteure aufgrund von Informationsdefiziten über die Leistungen des Einzelnen auf Informationen zurückgreifen, die über eine Gruppe vorliegen. Askriptive Merkmale dienen in diesem Fall für die Zuordnung der Individuen zu der jeweiligen Gruppe. Weichen die auf Gruppenebene zugeschriebenen durchschnittlichen Leistungen systematisch vom wahren Wert ab, so wird von Fehlerdiskriminierung (England 1992, S. 60) gesprochen. Im Anwendungsfall liegt eine Fehlerdiskriminierung beispielsweise dann vor, wenn die Leistungserwartungen, die an Mädchen im mathematischen Bereich oder an Jungen im sprachlichen Bereich gerichtet werden, von den tatsächlichen bereichsspezifischen Leistungen der jeweiligen Gruppe systematisch abweichen.

Allerdings lässt dieser Erklärungsansatz offen, wie und unter welchen Bedingungen es zu Fehlern im Rahmen der Schätzung von Gruppenmittelwerten kommt. Zur Beantwortung dieser Frage lassen sich (sozial-)psychologische Theorieansätze heranziehen. Dort werden zwei grundsätzliche Arten von Informationsquellen für die Einschätzung von Interaktionspartnern unterschieden (Brewer 1988). Neben eigenschaftsbasierten Wahrnehmungen anhand individueller verbaler und nonverbaler Merkmale existieren kategorisierende Wahrnehmungsprozesse. Letztere beinhalten den Rückgriff auf Stereotype. Dabei schreiben Individuen anderen Personen, die sie bestimmten Gruppen zuordnen, die Eigenschaften der Gruppe zu. Stereotype selbst sind als kognitive Abbilder oder Erwartungen über die Merkmale gesellschaftlicher Gruppen zu verstehen (Fiske 1998). Auf *individueller Ebene* bringen kategorisierende Wahrnehmungsprozesse entsprechend ein höheres Risiko der Fehleinschätzung mit sich als Einschätzungen anhand von merkmalsbasierten Wahrnehmungen, die eher zutreffend sind. Von Fehlern sind all diejenigen Individuen betroffen, die nicht dem Stereotyp entsprechen, anhand dessen sie eingeschätzt werden.

Die psychologische Forschung geht weiterhin davon aus, dass eigenschaftsbasierte und kategorisierende Informationsverarbeitungsprozesse die Pole eines Kontinuums

aus Verarbeitungsstilen bilden. Kategorisierungen erfolgen dem Kontinuum-Modell gemäß automatisch und gehen der kontrollierten und eigenschaftsbasierten Informationsverarbeitung voraus (Fiske et al. 1999). Ein Wechsel zu eher eigenschaftsbasierten Verarbeitungsstilen findet nur unter bestimmten Bedingungen statt. Neben der Verfügbarkeit von entsprechenden Informationen über die einzuschätzende Person müssen kategorien-inkonsistente Attribute vorliegen (Fiske 1998; Florack et al. 2001; Smith und DeCoster 2000). Da Grundschullehrkräfte die frisch eingeschulten Kinder zu Beginn des Schuljahres in der Regel noch nicht gut kennen, sollten schulleistungsrelevante Merkmale nur in einem geringen Umfang sichtbar sein. Aus diesem Grund dürften kategorisierende Verarbeitungsprozesse und Zuschreibungen, die auf Stereotypen basieren, zu Schuljahresbeginn eine besondere Rolle spielen; gleichzeitig sollten merkmalsbasierte Informationsverarbeitungsprozesse tendenziell weniger relevant sein als zu einem späteren Zeitpunkt im Verlauf des Schuljahres. Entsprechend sollten Leistungserwartungen von Lehrkräften zu frühen Zeitpunkten in den Schulkarrieren der Kinder stets in einem gewissen Umfang durch ethnische, soziale und geschlechtsspezifische Stereotype geprägt sein.

Da es sich bei Stereotypen um Generalisierungen handelt, die entweder in stärkerem Maße auf gesellschaftlich geteiltem Wissen oder in stärkerem Maße auf eigenen Erfahrungen im Umgang mit den betreffenden Gruppen basieren, können sie sowohl zutreffend als auch unzutreffend sein. Je nachdem, welche Stereotype über die jeweilige Gruppe vorliegen, sind deshalb auf der Grundlage stereotyper Informationsverarbeitungsprozesse sowohl negativ oder positiv verzerrte als auch zutreffende Erwartungen in Bezug auf die mittleren Leistungen verschiedener Schülergruppen möglich.

Allerdings bergen Stereotype das Risiko der fehlerhaften Vorhersage von Gruppeneigenschaften. Dies liegt insbesondere daran, dass Stereotype erfahrungsresistent sind und auch bei widersprüchlichen Informationen nicht ohne weiteres im Sinne einer zutreffenden Anpassung korrigiert werden (Sherman et al. 2005; Snyder und Stukas 1999). Stereotype sollten daher immer in einem gewissen Maße unzutreffend sein. Außerdem kann ihre Anwendung mit einer Akzentuierung tatsächlicher Gruppenunterschiede einhergehen. Entsprechend sollten Lehrererwartungen zu frühen Zeitpunkten in den Schulkarrieren auch auf *Gruppenebene* durch Fehler geprägt und deshalb systematisch verzerrt sein.

Im internationalen Kontext wurden in der Vergangenheit Stereotype nachgewiesen, die einigen Zuwanderergruppen und Kindern aus sozioökonomisch schwachen Familien, ebenso wie Jungen und Mädchen in bestimmten Kompetenzbereichen geringere Bildungserfolge zuschreiben (Arnold 2001, S. 75). Für die Bundesrepublik liegen kaum Studien vor, die auf lehrerseitige Stereotype hinweisen. Da Stereotype teilweise gesellschaftlich geteiltes Wissen widerspiegeln, sollte denjenigen Schülergruppen, für die in einem öffentlichkeitswirksamen Rahmen geringe schulische Erfolge festgestellt wurden, auch geringere und zum Teil negativ verzerrte Erwartungen entgegengebracht werden. Orientiert man sich an den Befunden der PISA-Studien (z. B. Prenzel und Baumert 2004), so sollte dies in erster Linie Schüler mit einem türkischen Zuwanderungshintergrund betreffen. Hinweise für negativ verzerrte Einschätzungen der Leistungen dieser Herkunftsgruppe liegen auf experimenteller Basis vor (Glock und Krolak-Schwerdt 2013; Sprietsma 2013).

Gleichzeitig sollten an Kinder mit Eltern oder Großeltern aus Osteuropa und der ehemaligen Sowjetunion, denen im Rahmen von PISA im Vergleich zur türkischstämmigen Gruppe geringere Nachteile bescheinigt wurden (Segeritz et al. 2010), im Schnitt weniger negativ verzerrte Erwartungen gerichtet werden als an die türkischstämmigen Kinder. Die Zugehörigkeit zu weniger privilegierten sozioökonomischen Schichten könnte auf die gleiche Weise zu negativ verzerrten Lehrererwartungen beitragen. Ähnliches gilt für die Wahrnehmung der Kompetenzen von Mädchen in mathematischen und Jungen in sprachlichen Fachbereichen.

3 Empirische Befundlage zu systematischen Verzerrungen in Lehrererwartungen

In der Literatur werden Leistungsurteile von Lehrkräften meist dann als unzutreffend deklariert, wenn eine Diskrepanz zwischen testbasierten Leistungsmessungen und lehrerbasierten Einschätzungen vorliegt (Ready und Wright 2011, S. 338 f.). In statistischen Modellen werden dazu die Effekte von Variablen zu den Leistungen, welche die „wahren“ Kompetenzen der Kinder abbilden sollen, auf Variablen zu Lehrererwartungen geschätzt. In einem zweiten Schritt werden zusätzlich askriptive Variablen wie das Geschlecht, die soziale oder die ethnische Herkunft in die Modelle aufgenommen. Weisen die askriptiven Merkmale einen signifikanten Effekt auf die Lehrereinschätzungen auf, der über die objektiv gemessenen Leistungen hinausgeht, so ist in der Literatur von systematisch verzerrten Lehrererwartungen die Rede.

Entsprechend dieses Vorgehens zeigen einige Arbeiten, dass Lehrkräfte negativ verzerrte Erwartungen an die schulischen Kompetenzen von Kindern haben, die ethnischen Minderheiten angehören oder die aus Zuwandererfamilien stammen (Baron et al. 1985; Jussim und Harber 2005; Jussim et al. 1996; Madon et al. 1998; McKown und Weinstein 2008; Ready und Wright 2011; Tenenbaum und Ruck 2007). Auch Schüler aus niedrigeren sozialen Schichten werden häufig unterschätzt (Alvidrez und Weinstein 1999; Baron et al. 1985; Dusek und Joseph 1983; Jussim et al. 1996; Madon et al. 1998). Diese beiden Herkunftsmerkmale (sozial, ethnisch) scheinen konfundiert auf die Entstehung von Lehrererwartungen zu wirken (Jussim et al. 1996, S. 313 ff.).

Weniger eindeutig sind dagegen die geschlechtsspezifischen Befunde. In manchen Untersuchungen werden die Fähigkeiten von Jungen trotz gleicher Leistungen niedriger eingeschätzt als die von Mädchen (Ehrenberg et al. 1995; Hartley 1982; Jussim 1989; Jussim und Eccles 1995; Ross und Jackson 1991), während in anderen Studien die Fähigkeiten von Jungen überschätzt werden (Doherty und Conolly 1985; Eccles und Blumenfeld 1985; Parsons et al. 1982). Dies könnte damit zusammenhängen, dass geschlechtsspezifische Verzerrungen in Lehrererwartungen bereichsspezifisch sind. Überschätzungen der Fähigkeiten von Jungen treten eher im mathematischen Bereich auf (Jussim und Eccles 1992; McKown und Weinstein 2002), während Mädchen eher im Bereich Lesen überschätzt werden (Hinnant et al. 2009). Gleichzeitig wird den Mädchen ein höheres Maß an Fleiß zugeschrieben, während bei Jungen mehr Talent vermutet wird (Jussim 1989; Jussim und Eccles 1992). Allerdings finden sich auch Untersuchungen, die keine geschlechtsspezifisch

geprägten Einschätzungen nachweisen (Dusek und Joseph 1983; Hoge und Butcher 1984; Hoge und Coladarci 1989).

Trotz eines weitgehend ähnlichen Vorgehens beim empirischen Nachweis verzerrter Lehrererwartungen erweist sich der Forschungsstand als gemischt. Zum Teil lassen sich verzerrte Erwartungen nachweisen, in einigen Fällen jedoch nicht (vgl. für eine Zusammenschau Jussim und Harber 2005; Jussim et al. 2009; Schofield 2006; Spitz 1999). Diese Inkonsistenz könnte mit den jeweils eingesetzten Messungen der Erwartungen zusammenhängen. Nur die wenigsten nicht-experimentellen Studien können hierbei auf Indikatoren zurückgreifen, die auf einer tatsächlichen Messung von Erwartungen im Sinne lehrerseitiger Prognosen basieren. Stattdessen werden meistens Beurteilungen einer oder mehrerer Kompetenzen oder des Talents – bezogen auf den Ist-Zustand – analysiert. Ob sich systematische Verzerrungen auch dann finden lassen, wenn die Erwartungen explizit auf zukünftige Schülerleistungen bezogen werden, ist nicht bekannt.

Eine weiteres Problem betrifft den Zeitpunkt der Messung. In fast allen Arbeiten werden Lehrerurteile zu einem Zeitpunkt erhoben, zu dem die Lehrkräfte ihre Schülerschaft bereits seit längerer Zeit kennen. So fanden die Erhebungen der „Michigan Study of Adolescent Life Transitions (MSALT)“, aus der zahlreiche wegweisende Arbeiten zum Thema hervorgingen (Jussim 1989; Jussim und Eccles 1992; Jussim et al. 1996; Madon et al. 1997, 1998), erst im sechsten Schuljahr statt. Zu diesem Zeitpunkt hatten die Lehrkräfte die Kinder schon über einen längeren Zeitraum unterrichtet. Die in der Studie gemessenen Kompetenzen könnten bereits durch vorausgegangene selbsterfüllende Prophezeiungen beeinflusst worden sein. Ebenso könnten die Lehrkräfte ihre Erwartungen an die beobachteten Schülermerkmale angepasst haben. In welche Richtung Anpassungen stattgefunden haben, lässt sich ausgehend von den verfügbaren Messungen nicht feststellen.

Ein Großteil dieser Forschung stammt aus den USA. Im deutschen Kontext werden Lehrererwartungen zwar angesprochen (z. B. Dumke 1977; Friedrich et al. 2015; Hirschauer und Kullmann 2010; Ludwig und Ludwig 2007; Schofield 2006; Ulich 2001), einschlägige empirische Arbeiten sind jedoch die Ausnahme. Lediglich zwei Studien im deutschen Kontext liefern Anhaltspunkte für systematische Verzerrungen. So deuten die Ergebnisse von Stahl (2007) darauf hin, dass die Lesefähigkeit von Jungen und Kindern aus sozial schwächeren Familien unterschätzt werden. Darüber hinaus lässt sich ein Einfluss des Zuwanderungshintergrunds unter Kontrolle der sozialen Herkunft auf die Einschätzung der allgemeinen schulischen Begabung feststellen. Die Erfassung des Zuwanderungshintergrunds über die Herkunftssprache (Deutsch vs. nicht Deutsch) lässt allerdings keine Aussagen über möglicherweise bestehende Unterschiede in den Lehrerurteilen gegenüber verschiedenen ethnischen Herkunftsgruppen zu. Die Untersuchung der eingeschätzten Begabung sagt zudem nichts über lehrerseitige Erwartungen an die bereichsspezifischen Leistungen der Schüler aus, beispielsweise im Lesen oder im mathematischen Bereich.

In einer zweiten Studie aus Deutschland stehen Lehrererwartungen an die Wahrscheinlichkeit das Abitur zu erreichen im Mittelpunkt (Becker 2013). Für Jugendliche der zehnten Klassenstufe wurde gezeigt, dass diese Wahrscheinlichkeit für Schüler aus sozioökonomisch besser gestellten Familien bei gleichen Leistungen überschätzt wird. Weitere askriptive Schülermerkmale wurden in der Untersuchung jedoch nicht

berücksichtigt. Kritisch anzumerken ist außerdem, dass die der Studie zugrunde liegenden Daten aus dem Jahr 1969 stammen und sich die Ergebnisse deshalb nur bedingt auf aktuelle Kohorten von Schülern oder Lehrkräften übertragen lassen.

In der vorliegenden Studie wird daher untersucht, ob Lehrpersonen Kindern unterschiedlichen sozialen und ethnischen Hintergrunds und unterschiedlichen Geschlechts zu Beginn der ersten Grundschulklasse verschieden hohe Erwartungen entgegenbringen und inwiefern diese Unterschiede tatsächliche Differenzen in den schulischen Kompetenzen zwischen den Schülergruppen widerspiegeln.

Auf Grundlage der theoretischen Überlegungen und des Forschungsstands wird erwartet, dass Lehrererwartungen an die zukünftigen schulischen Leistungen von Kindern aus Familien mit einem niedrigeren sozioökonomischen Status geringer sind. Außerdem sollten Lehrkräfte geringere Erwartungen an die Leistungen von Kindern türkischer Herkunft im Vergleich zu Kindern ohne Zuwanderungshintergrund und im Vergleich zu osteuropäischen Herkunftsgruppen haben. Hinsichtlich des Geschlechts werden den gängigen Geschlechtsstereotypen entsprechende Unterschiede vermutet, d. h. im mathematischen Bereich geringere Erwartungen an Mädchen und im sprachlichen Bereich geringere Erwartungen an Jungen. Ob diese Unterschiede über die tatsächlichen Fähigkeiten und Leistungen der Kinder hinaus bestehen bleiben und systematische Verzerrungen widerspiegeln, hängt davon ab, ob die zugrunde liegenden Stereotype unzutreffend sind. Dabei handelt es sich um eine empirisch zu klärende Frage.

4 Methodik

4.1 Datengrundlage

Im Schuljahr 2013/2014 nahmen 1065 Kinder aus 68 ersten Klassen und 77 Klassen- und Fachlehrkräfte, die in diesen Grundschulklassen unterrichteten, an der Untersuchung „Kompetenzerwerb und Lernvoraussetzungen“ (KuL) teil.¹ Die Erhebungen fanden in verschiedenen Städten des Ruhrgebiets (Nordrhein-Westfalen) statt. Vor Beginn der Erhebungen wurden 409 Schulen kontaktiert, von denen sich 39 Grundschulen beteiligten.² Im Falle einer positiven Antwort wurden alle Klassenlehrkräfte, die im Schuljahr 2013/2014 die ersten Klassen der Schule leiten sollten, um ihr Einverständnis zur Teilnahme an der Studie gebeten. Hier beläuft sich der Rücklauf auf 74 %. Über die teilnehmenden Lehrkräfte wurden unmittelbar zu Beginn des Schuljahres Anschreiben an alle Kinder der jeweiligen Klasse verteilt. Darin wurden die Eltern gebeten, der Teilnahme ihres Kindes an der Studie zuzustimmen. Die Teilnahmequote auf Eltern-Kind-Ebene lag bei 66 %.

¹ Dieses Forschungsprojekt wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01JC1117 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

² Die geringe Ausschöpfung auf Schulebene (16,4 %) könnte mit der mittlerweile stark gestiegenen Anzahl an Feldstudien im Bereich der Bildungsforschung zusammenhängen. In der Mehrzahl der nicht teilnehmenden Schulen gaben die Schulleitungen als Grund ihrer Absage eine zu hohe Belastung der Lehrerschaft durch zurzeit laufende oder bereits abgeschlossene wissenschaftliche Studien an.

Die Lehrkräfte wurden unmittelbar nach Schuljahresanfang zu ihren Erwartungen an die schulleistungsbezogene Entwicklung eines jeden teilnehmenden Kindes befragt. Dabei wurde auf den erwarteten Verlauf des soeben begonnenen Schuljahres Bezug genommen. Zeitgleich zur Befragung der Lehrkräfte wurden die sprachlichen und mathematischen Fähigkeiten der Kinder getestet. Um Informationen zum familiären Hintergrund und zur häuslichen Lernumwelt zu erheben, fanden im September und Oktober 2013 außerdem telefonische Interviews mit den Eltern statt. Ein Teil der Eltern konnte erst im Juni/Juli 2014 im Rahmen einer schriftlichen Nachbefragung erreicht werden.

Diese Datengrundlage lässt Aussagen über Lehrererwartungen zu, die so bisher weder im deutschen noch im internationalen Kontext möglich waren. Dies liegt einerseits am minimalen Lehrkraft-Kind-Kontakt, der den Erhebungen vorausgegangen ist, und andererseits an der gleichzeitigen Durchführung von Leistungstests und Lehrerbefragung. Ebenfalls hervorzuheben ist, dass die Messung der Lehrererwartungen nicht wie in den meisten vorliegenden Studien über die Abfrage des Ist-Zustandes, sondern als Einschätzung der zukünftigen Entwicklung des Kindes erfolgte. Außerdem lassen sich ausgehend von der Erfassung spezifischer Erwartungen zu den Unterrichtsfächern Deutsch und Mathematik verschiedene Kompetenzbereiche miteinander vergleichen.

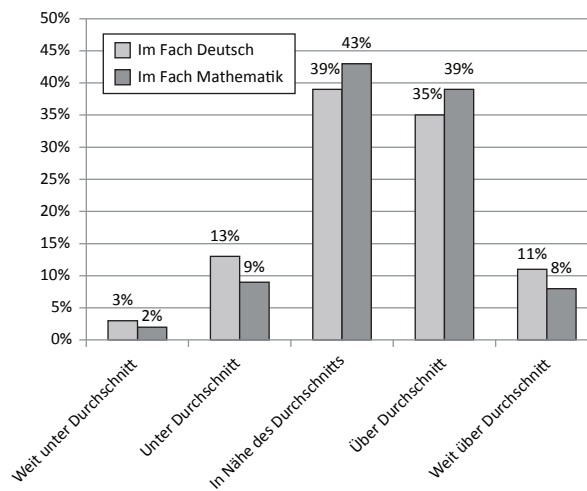
4.2 Instrumente und deskriptive Verteilungen

Die Kinder der Stichprobe waren zum Zeitpunkt der Einschulung im Mittel 6 Jahre und 5 Monate alt ($SD=0,4$; $Min=5,7$; $Max=8,0$). Die Anzahl teilnehmender Kinder je Klasse lag bei durchschnittlich 12 Kindern. Bei den Lehrkräften handelte es sich, wie für Grundschulen in Deutschland üblich, vorwiegend um weibliche Lehrpersonen (95 %) ohne Zuwanderungshintergrund (97 %). Die Lehrkräfte waren zu Beginn des Schuljahres 2013/2014 im Durchschnitt 41 Jahre und 5 Monate ($SD=8,9$; $Min=26,8$; $Max=61,3$) alt und seit durchschnittlich 12 Jahren ($SD=8,85$) im Lehrberuf tätig.

Zur Messung von *Lehrererwartungen* schätzten die Lehrkräfte am Schuljahresbeginn auf einer fünfstufigen Skala ein, wo das Kind mit seinen Leistungen am Ende des kommenden Schuljahres im Vergleich zu den anderen Kindern aus seiner Klasse voraussichtlich stehen wird (1 = weit unter dem Klassendurchschnitt, 2 = unter dem Klassendurchschnitt, 3 = in der Nähe des Klassendurchschnitts, 4 = über dem Klassendurchschnitt, 5 = weit über dem Klassendurchschnitt).

Wie Abb. 1 zeigt, nutzten die Lehrkräfte für ihre Einschätzung zur Entwicklung der Kinder sowohl im Fach Deutsch als auch Mathematik die gesamte Skalenbreite aus. Die Tendenz geht jedoch zu mittleren und leicht überdurchschnittlichen Erwartungen, sodass die Antwortkategorien „erwarteter Leistungsstand in der Nähe des Durchschnitts“ und „erwarteter Leistungsstand über dem Durchschnitt“ mit jeweils etwa 40 % am häufigsten vorkommen. Die Einschätzungen für beide Fächer weisen eine bedeutsame Varianz auf Individualebene auf (Deutsch: $\sigma_w=0,88$, $S.E.=0,05$, $p \leq 0,001$; Mathematik: $\sigma_w=0,71$, $S.E.=0,05$, $p \leq 0,001$), wobei kaum Variationen zwischen den Klassen zu verzeichnen sind (Deutsch: $\sigma_B=0,01$, $S.E.=0,01$, n.s.; Mathematik: $\sigma_B=0,00$, $S.E.=0,01$, n.s.). Entsprechend klein fallen auch die Intra-

Abb. 1: Häufigkeitsverteilung der Lehrererwartungen in Bezug auf die Leistungen in den Fächern Deutsch und Mathematik



klassenkorrelationen aus, die den Anteil der Varianz zwischen den Klassen an der Gesamtvarianz kennzeichnen (Deutsch: $ICC=0,003$, Mathematik: $ICC=0,001$). Variationen der Lehrererwartungen sind damit vorwiegend zwischen verschiedenen Kindern zu finden und lassen sich nur zu einem minimalen Anteil auf Unterschiede im Antwortverhalten zwischen verschiedenen Lehrkräften zurückführen.

Zur Bestimmung des *Zuwanderungshintergrunds* wurden Elternangaben zum Geburtsland des Kindes, beider Elternteile sowie der Großeltern verwendet. Ein Zuwanderungshintergrund liegt vor, wenn mindestens ein (Groß-) Elternteil nicht in Deutschland geboren wurde. Die Kinder der ersten Zuwanderergeneration (21 Fälle) wurden nicht berücksichtigt. Dasselbe trifft auf Kinder ohne Angaben zum Zuwanderungshintergrund zu (134 Fälle). Die sich daraus ergebende Analysestichprobe ($N=901$) erlaubte die separate Betrachtung der Herkunftsländer „Türkei“ (11%), „Osteuropa“ (11%) und „sonstige Länder“ (16%), wobei die Kategorie „Osteuropa“ alle Staaten der ehemaligen Sowjetunion sowie Polen, Tschechien, Bulgarien, Rumänien, Moldawien, Ungarn, Slowakei und Slowenien umfasst (vgl. Tab. 1).³ Die Referenzgruppe bilden Kinder ohne Zuwanderungshintergrund (62%). Die Zuordnung zu den verschiedenen Herkunftsgruppen entspricht dem Vorgehen des Nationalen Bildungspanels (Olczyk et al. 2014).

Als Indikator des *sozioökonomischen Status* wurde der „HISEI“ (Highest International Socio-Economic Index of Occupational Status; Ganzeboom et al. 1992) herangezogen. Dieser wurde basierend auf der „ISCO-08“ (International Standard Classification of Occupations; Ganzeboom 2010) bestimmt. Tabelle 1 verdeutlicht, dass sich die Herkunftsgruppen deutlich in ihrem sozialen Status unterscheiden. Die höchsten HISEI-Werte sind in der Gruppe ohne Zuwanderungshintergrund zu finden, während Kinder türkischer Herkunft und aus anderen Ländern im Schnitt den sozioökonomisch schwächsten Schichten angehören. Die Gruppe mit osteuropäischem Zuwanderungshintergrund liegt dazwischen.

³Die Anzahl der in die Analysen eingehenden Lehrkräfte verringerte sich durch diese Einschränkungen auf 69.

Tab. 1: Deskriptive Statistik der Analysevariablen

	Mittelwert	SD	Min	Max	Fehlend
<i>Lehrererwartungen</i>					
Im Fach Deutsch	3,41	0,94	1	5	28
Im Fach Mathematik	3,37	0,84	1	5	62
<i>Leistung</i>					
Lautbewusstheit	13,54	4,57	1	26	41
Lesen	23,99	26,99	0	156	41
Mathematik	27,58	7,73	3	53	41
<i>Allg. kognitive Fähigkeiten</i>					
Matrizentest	6,73	2,88	0	12	34
Kurzzeitgedächtnis	3,18	1,67	0	8	41
<i>Motivationale Merkmale</i>					
Lernfreude	1,31	0,42	0	2	36
Anstrengungsbereitschaft	1,48	0,40	0	2	35
HISEI	53,25	19,48	14	89	14
Deutsch	57,17	18,76	15	89	7
T9,76chg	43,11	18,23	14	89	0
Osteuropäisch	49,40	18,46	14	87	1
Anderes Land	47,96	19,55	14	87	6
	Anzahl	Prozent			
<i>Ethnischer Hintergrund</i>					
Deutsch	555	61,6%			0
Türkisch	102	11,3%			0
Osteuropäisch	98	10,9%			0
Anderes Land	146	16,2%			0
<i>Geschlecht</i>					
Weiblich	429	48,0%			0
Männlich	472	52,0%			0

HISEI highest international socio-economic index of occupational status. *N*=901

Als dritte soziale Gruppierung diente das *Geschlecht des Kindes* (0=Männlich, 1=Weiblich). Die Verteilung zeigt einen etwas geringeren Anteil an Mädchen (48%) als Jungen (52%) in der Stichprobe.

Mathematische und sprachliche Fähigkeiten der Kinder wurden anhand des Instruments „Fähigkeitsindikatoren Primarschule“ (FIPS; Bäuerlein et al. 2012) getestet. In die Analysen gingen die Skalen „Mathematik“ (53 Items), „Lautbewusstheit“ (26 Items) und „Lesen“ (159 Items) ein. Als Maße der *allgemeinen kognitiven Fähigkeiten*, welche die Grundfähigkeiten des schlussfolgernden Denkens sowie der Arbeitsgedächtnisleistung abbilden sollen, dienten der Kurzzeitgedächtnistest des FIPS und der Subtest „Matrizen“ aus dem Grundintelligenztest Skala 1 (CFT 1; Weiß und Osterland 1997). Alle Leistungsvariablen gingen als am Gesamtmittelwert zentrierte Summenwerte in die Analysen ein.

Motivationale Merkmale der Kinder wurden anhand einer gekürzten und adaptierten Version der Skalen „Lernfreude“ (5 Items) und „Anstrengungsbereitschaft“ (5 Items) aus dem Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen (FEESS 1–2, Rauer und Schuck 2004) von den Kindern erfragt. Die Items der Lernfreude und Anstrengungsbereitschaft

wurden zu Skalenmittelwerten zusammengefasst und ebenfalls am Gesamtmittelwert zentriert.

4.3 Statistische Analysen

Um Verzerrungen in den Lehrerwartungen zu untersuchen, wurden Random-Intercept-Modelle der Lehrererwartung auf mathematische und sprachliche sowie allgemeine kognitive Fähigkeiten, den Zuwanderungshintergrund, den sozioökonomischen Status der Familien sowie auf das Geschlecht der Kinder berechnet.⁴ Die Schätzungen der statistischen Kennwerte wurden mit der Full Information Maximum-Likelihood Methode (FIML, vgl. Schafer und Graham 2002) unter Anwendung der Software Mplus Version 7.3 (Muthén und Muthén 2012) durchgeführt. FIML schließt Fälle mit fehlenden Werten in einer oder mehreren unabhängigen Variablen in die Schätzungen ein und ersetzt diese durch den Rückgriff auf Informationen, die durch zusätzliche, nicht in die Analysen eingehende Hilfsvariablen bereitgestellt werden.

5 Ergebnisse

Um Aussagen darüber treffen zu können, ob Lehrererwartungen mehr oder weniger zutreffend sind, wurde in einem ersten Schritt untersucht, inwieweit sich die tatsächlichen Leistungen der Kinder in den Erwartungen der Lehrkräfte niederschlagen (Tab. 2). Sowohl für die Lehrererwartungen im Fach Deutsch als auch im Fach Mathematik zeigen die Analysen, dass sich über die Kompetenzen, die kognitiven Fähigkeiten und die Motivation der Kinder 36,7% (Deutsch) und 35,6% (Mathematik) der Varianz auf Individualebene in den Lehrererwartungen aufklären lässt. Dieser Anteil der Unterschiede in den Lehrererwartungen kann als zutreffend erachtet werden. Gleichzeitig bleibt ein bedeutsamer und deutlich größerer Anteil unerklärt. Die entscheidende Frage für die vorliegende Untersuchung ist, inwieweit askriptive Schülermerkmale diese Varianz aufklären können.

Zunächst soll geklärt werden, ob sich die Lehrererwartungen gegenüber verschiedenen Zuwanderergruppen unterscheiden (Tab. 3). Die Ergebnisse hierzu fallen in den Fächern Deutsch und Mathematik insgesamt ähnlich aus (Modell 1). In beiden Fachbereichen erwarten Lehrkräfte für Kinder mit türkischem Zuwanderungshintergrund im Vergleich zu Kindern ohne Zuwanderungshintergrund weniger gute Leistungen. Dagegen erweisen sich die Erwartungen für Kinder osteuropäischer Herkunft in beiden Unterrichtsfächern im Vergleich zu Kindern ohne Zuwanderungshintergrund als höher.

Diese ersten Analysen weisen darauf hin, dass sich die Erwartungen von Grundschullehrkräften für zukünftige Schülerleistungen zu Beginn des ersten Schuljahres systematisch in Abhängigkeit vom ethnischen Hintergrund der Schüler unterscheiden.

⁴Die Analysemodelle enthalten die beiden Ebenen der Kinder und der Lehrkräfte. Auf diese Weise können Unterschiede zwischen verschiedenen Lehrkräften im Antwortverhalten berücksichtigt werden (vgl. Snijders und Bosker 1999). Dies ist bedeutsam, weil die Lehrererwartungen im Vergleich zu den anderen Kindern aus der Klasse gemessen wurden. Aus diesem Grund könnten die Einschätzungen innerhalb von Lehrkräften ähnlicher sein als zwischen den Lehrkräften.

Tab. 2: Vorhersage von Lehrererwartungen in Bezug auf die Leistungen in den Fächern Deutsch und Mathematik

	Deutsch			Mathematik		
	b	SE b	β	B	SE b	β
<i>Leistung</i>						
Lautbewusstheit	0,06	0,01	0,30***			
Lesen	0,01	0,00	0,24***			
Mathematik				0,05	0,00	0,45***
<i>Allg. kog. Fähigkeiten</i>						
Matrizentest	0,04	0,01	0,13***	0,05	0,01	0,17***
Kurzzeitgedächtnis	0,06	0,01	0,10***	0,02	0,01	0,03
<i>Motivationale Merkmale</i>						
Lernfreude	0,11	0,07	0,05	0,01	0,05	0,01
Anstrengungsbereitschaft	0,28	0,08	0,12***	0,25	0,08	0,12**
Within-Residualvarianz	0,55***	0,03		0,45***	0,03	
Between-Residualvarianz	0,04*	0,02		0,02*	0,01	
Intercept	3,38***	0,04		3,41***	0,03	
Within- R^2	36,7%***			35,6%***		

Random-Intercept-Modelle der Lehrererwartungen auf die Leistungen in den Unterrichtsfächern Deutsch und Mathematik. $N=895$

* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$

Offen bleibt allerdings die Frage, inwiefern die beobachteten Ungleichheiten tatsächliche Leistungsunterschiede widerspiegeln. Um dies zu untersuchen, wurden jeweils bereichskongruente Leistungsmerkmale sowie allgemeine kognitive Fähigkeiten und die Motivation der Kinder in die Modelle aufgenommen (Modell 2).

Zuerst werden die Ergebnisse zum Fach Deutsch betrachtet. Als kongruente Leistungsmerkmale sind hier die Testergebnisse in den FIPS-Skalen „Lautbewusstheit“ und „Lesen“ kontrolliert. Auch bei gleicher Leistung erwarten Lehrkräfte von türkisch-stämmigen Kindern eine ungünstigere Leistungsentwicklung als von einheimischen Kindern. Diese Unterschiede in den Lehrererwartungen gehen nicht auf die gemessenen Leistungsindikatoren zurück, sondern könnten stereotype Verzerrungen zum Nachteil türkisch-stämmiger Kinder kennzeichnen. Der Vorteil in den Lehrererwartungen zugunsten von Kindern mit einem osteuropäischen Zuwanderungshintergrund bleibt unter Kontrolle der Leistungsmerkmale dagegen nicht bestehen. Die höheren Erwartungen an diese Schülergruppe im Vergleich zu Kindern aus der Mehrheitsbevölkerung scheinen also eine zutreffende Einschätzung widerzuspiegeln.

Anders stellen sich die Ergebnisse im Fach Mathematik unter Kontrolle der Mathematikleistung, der allgemeinen kognitiven Fähigkeiten und der Motivation der Kinder dar. Ein Nachteil bei den türkisch-stämmigen Kindern ist hier nicht mehr sichtbar. In Bezug auf die zukünftigen Leistungen im Fach Mathematik scheinen die Lehrererwartungen an diese Schülergruppe also nicht verzerrt zu sein. Allerdings ist ersichtlich, dass Lehrkräfte bei gleicher Mathematiktestleistung und gleichen allgemeinen kognitiven Fähigkeiten von Kindern osteuropäischer Herkunft signifikant höhere Mathematikleistungen als von Kindern ohne Zuwanderungshintergrund erwarten. Die Erwartungen an diese Schülergruppe scheinen damit positiv verzerrt zu sein.

Tab. 3: Prädiktoren der Lehrererwartungen in Bezug auf die Leistungen in den Unterrichtsfächern Deutsch und Mathematik

	Modell 1						Modell 2					
	Deutsch			Mathematik			Deutsch			Mathematik		
	b	SE b	β	b	SE b	β	b	SE b	β	b	SE b	β
<i>Ethnischer Hintergrund</i>												
Türkisch	-0,28	0,09	-0,10***	-0,18	0,07	-0,07**	-0,28	0,09	-0,10**	-0,10	0,06	-0,04
Osteuropäisch	0,23	0,10	0,08*	0,35	0,10	0,13***	0,09	0,08	0,03	0,24	0,09	0,09**
Anderes Land	-0,20	0,09	-0,08*	-0,06	0,08	-0,03	-0,24	0,08	-0,10**	0,00	0,08	0,00
<i>Sozioökonomischer Status</i>												
HISEI	0,01	0,00	0,27***	0,01	0,00	0,26***	0,01	0,00	0,16***	0,01	0,00	0,15***
<i>Geschlecht</i>												
Weiblich	0,22	0,08	0,12**	-0,13	0,06	-0,08*	0,17	0,06	0,09**	-0,02	0,06	-0,01
<i>Leistung</i>												
Lautbewusstheit							0,05	0,01	0,26***			
Lesen							0,01	0,00	0,25***	0,04	0,00	0,41***
Mathematik												
<i>Allg. kognitive Fähigkeiten</i>												
Matrizentest							0,03	0,01	0,10**	0,05	0,01	0,16***
Kurzzeitgedächtnis							0,07	0,01	0,13***	0,03	0,01	0,05*
<i>Motivationale Merkmale</i>												
Lernfreude							0,12	0,08	0,05	0,03	0,05	0,02
Anstrengungsbereitschaft							0,26	0,07	0,11***	0,22	0,08	0,11**
Within-Residualvarianz	0,78***	0,05		0,64***	0,04		0,51***	0,03		0,43***	0,03	
Between-Residualvarianz	0,00	0,01		0,00	0,01		0,05**	0,02		0,03*	0,01	
Intercept	3,30***	0,05		3,47***	0,04		3,36***	0,05		3,41***	0,04	
Within- R^2	12,0%***			10,1%***			40,9%***			37,8%***		

Random-Intercept-Modelle der Lehrererwartungen in Bezug auf die Leistungen in den Unterrichtsfächern Deutsch und Mathematik. $N=901$

HISEI highest international socio-economic index of occupational status

* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$

Diese Ergebnisse machen insgesamt deutlich, dass Lehrkräfte im Fach Deutsch systematisch negativ verzerrte Erwartungen an türkisch-stämmige Kinder richten, während die Leistungsentwicklungen von Kindern osteuropäischer Herkunft im Fach Mathematik zu hoch eingeschätzt zu werden scheinen. Betont werden muss, dass die beiden Herkunftsgruppen jeweils unabhängig von ihrer sozialen Herkunft im Vergleich zu Kindern ohne Zuwanderungshintergrund unter- bzw. überschätzt werden; denn die soziale Herkunft wird in den zugrundeliegenden Modellen kontrolliert. Es kann also explizit von ethnischen Verzerrungen die Rede sein.

Auch mit dem sozioökonomischen Status der Kinder sind unterschiedliche Lehrererwartungen verbunden (Modell 1). Allgemein werden höhere schulische Leistungen für Kinder aus sozioökonomisch besser gestellten Familien als für Kinder aus Familien mit niedrigerem sozialem Status erwartet. Die Effekte für die beiden Unterrichtsfächer Mathematik und Deutsch sind gleich stark ausgeprägt. Werden die tatsächlichen Leistungsunterschiede im jeweiligen Bereich sowie die kognitiven Fähigkeiten und die motivationalen Variablen kontrolliert, reduzieren sich die Effekte, bleiben aber sowohl für das Fach Deutsch als auch für das Fach Mathematik bedeutsam (Modell 2). In beiden Unterrichtsfächern scheinen die Lehrererwartungen damit systematisch mit der sozialen Herkunft der Schüler in Verbindung zu stehen, und zwar unabhängig davon, ob es sich um Mädchen oder Jungen handelt und ob ein Zuwanderungshintergrund vorliegt oder nicht.

Schließlich zeigt sich im Fach Deutsch ein deutlicher Vorteil in den Erwartungen an Mädchen im Vergleich zu Jungen (Modell 1). Dieser Unterschied bleibt auch dann bestehen, wenn die Leistungen in der Lautbewusstheit und der Lesekompetenz, die kognitiven Fähigkeiten sowie die Motivation kontrolliert werden (Modell 2). Hierin könnte sich möglicherweise eine Verzerrung der Lehrererwartungen zeigen, die dem Geschlechtsstereotyp, Mädchen seien im sprachlichen Bereich begabter als Jungen, entspricht.

Im Unterrichtsfach Mathematik finden sich dagegen keine geschlechtsspezifischen Verzerrungen. Die Lehrkräfte erwarten in diesem Kompetenzbereich zwar höhere Leistungen von Jungen im Vergleich zu Mädchen (Modell 1). Diese Erwartung scheint aber den tatsächlichen Kompetenzen der Kinder zu entsprechen (Modell 2).⁵

6 Fazit und Diskussion

In dem vorliegenden Beitrag wurde der Frage nachgegangen, ob Grundschullehrkräfte mit unterschiedlich hohen Erwartungen an verschiedene Schülergruppen herantreten und ob diese Unterschiede im Vergleich zu den tatsächlichen Leistungen und Fähig-

⁵Die Lehrkräfte wurden zwischen der dritten und der zwölften Unterrichtswoche zu ihren Erwartungen befragt. Um zu überprüfen, ob der Zeitpunkt der Datenerhebung das Ausmaß der Verzerrungen in den Lehrererwartungen beeinflusst, wurde in weiterführende Analysemodelle das Datum des Rückversands der Lehrkraftfragebögen in Interaktion mit den askriptiven Merkmalen der Kinder aufgenommen. Die Ergebnisse (nicht dargestellt) verdeutlichen, dass die in den beiden Fachbereichen gefundenen Verzerrungen in den Lehrererwartungen nicht in Abhängigkeit des Versanddatums der Fragebögen variieren. Außerdem entsprechen die Haupteffekte der ethnischen und der sozialen Herkunft sowie des Geschlechts der Kinder den in den Tab. 2 und 3 berichteten Ergebnissen.

keiten der Kinder korrekt eingeschätzt werden. Ausgehend von den theoretischen Überlegungen wurde erwartet, dass die lehrerseitige Wahrnehmung zum Teil durch die Aktivierung und Anwendung von unzutreffenden Stereotypen geprägt ist.

Die empirischen Analysen weisen insgesamt darauf hin, dass Lehrererwartungen nach askriptiven Schülermerkmalen verzerrt zu sein scheinen. Das heißt, dass die Erwartungen für Kinder unterschiedlicher sozialer und ethnischer Herkunft und unterschiedlichen Geschlechts auch unter Kontrolle tatsächlicher Kompetenzen, kognitiver Grundfähigkeiten und motivationaler Merkmale systematisch variieren.

Die Ergebnisse bestätigen internationale Arbeiten, die belegen, dass Lehrererwartungen zu einem gewissen Anteil unzutreffend sind. Zusätzlich zeigt sich, dass sich systematische Verzerrungen je nach betrachtetem Kompetenzbereich unterscheiden. Außerdem scheinen Verzerrungen nach der sozialen und ethnischen Herkunft und dem Geschlecht der Kinder unabhängig voneinander zu bestehen. Im Unterrichtsfach Deutsch finden sich negative Verzerrungen in den Erwartungen der Lehrkräfte an Kinder mit einem türkischen Zuwanderungshintergrund, an Kinder aus sozial schwächeren Familien sowie an Jungen. Bei der Betrachtung des Unterrichtsfachs Mathematik finden sich ebenfalls Verzerrungen nach der sozialen Herkunft. Abgesehen von der Überschätzung von Kindern mit einem osteuropäischen Zuwanderungshintergrund im Vergleich zu einheimischen Kindern sind allerdings keine Verzerrungen nach der ethnischen Herkunft der Kinder sichtbar. Das gleiche trifft auf Unterschiede zwischen Mädchen und Jungen zu.

Dass sich ethnische Verzerrungen in den Lehrererwartungen je nach Herkunftsgruppe unterscheiden, verweist auf gruppenspezifisch variierende ethnische Stereotype. Die Unterschätzung von türkisch-stämmigen Kindern im sprachlichen Bereich lässt außerdem vermuten, dass sich die bestehenden ethnischen Stereotype an sozial geteiltem Wissen orientieren. So könnten sich die Stereotype der Lehrkräfte aus Forschungsergebnissen speisen, die in den Medien verbreiteten wurden. Beispielsweise ergeben die Befunde aus PISA in Bezug auf Kinder mit einem türkischen Zuwanderungshintergrund ein Muster, welches in seiner Richtung den in der vorliegenden Studie belegten Lehrererwartungen entspricht.

Dagegen überrascht der Befund, dass Kinder mit einem osteuropäischen Zuwanderungshintergrund in Mathematik im Vergleich zu einheimischen Kindern positiv verzerrt wahrgenommen werden. Mit den PISA-Daten wurde zwar gezeigt, dass Kinder, deren Eltern in Ländern der ehemaligen Sowjetunion oder in Polen geboren sind, im Bereich Mathematik besser abschneiden als Kinder aus türkischen Familien (Segeritz et al. 2010). Hinweise auf Vorteile gegenüber Kindern ohne Zuwanderungshintergrund lassen sich dort aber nicht finden. Aus der gegenwärtig verfügbaren Literatur geht auch nicht hervor, welche Stereotype über diese Gruppe verbreitet sind, ob diese verzerrt sind und ob sie für die hier berichteten Einschätzungen sorgen könnten. Ebenso ist nicht bekannt, ob sich Stereotype über osteuropäische Zuwanderer überhaupt aus Ergebnissen wie denen der PISA-Studie speisen. Schließlich ist nicht auszuschließen, dass die osteuropäische Herkunftsgruppe in der betrachteten Stichprobe im Hinblick auf nicht untersuchte Merkmale, welche die Lehrererwartungen in der gefundenen Weise beeinflussen, selektiv ist.

Die Erwartungen an die Leistungen von Kindern unterschiedlicher sozialer Herkunft könnten an sozialen Stereotypen ausgerichtet sein. Die Richtung der Ver-

zerrungen deckt sich mit dem gängigen Stereotyp, dass Kinder aus sozial schwächeren Familien schlechtere schulische Leistungen erzielen als Kinder aus sozial besser gestellten Familien. Auch diese Ergebnisse entsprechen Befunden über die Bildungsergebnisse von Kindern unterschiedlicher sozialer Herkunft (z. B. Prenzel und Baumert 2004).

Die Ergebnisse deuten weiterhin darauf hin, dass geschlechtsspezifische Stereotype zutreffende Abbilder der Leistungen von Mädchen und Jungen im Fachbereich Mathematik darstellen. In Übereinstimmung mit vorliegenden Forschungsarbeiten, die eine Antizipation von leistungsbezogenen Vorteilen für Jungen im mathematischen Bereich nahelegen (vgl. Hannover 2004), werden die Kompetenzen von Jungen und Mädchen den vorliegenden Ergebnissen zufolge in Bezug auf das Unterrichtsfach Mathematik zwar unterschiedlich, aber nicht verzerrt wahrgenommen. Geschlechtsspezifische Verzerrungen finden sich allerdings im sprachlichen Bereich. Das Stereotyp „Mädchen sind sprachlich talentierter als Jungen“ scheint demzufolge weniger zutreffend zu sein, als das Stereotyp „Jungen sind mathematisch talentierter als Mädchen“.

Zusammenfassend lässt sich festhalten, dass in beiden beobachteten Fachbereichen systematische Verzerrungen der Lehrererwartungen nach askriptiven Schülermerkmalen beobachtbar sind. Diese spiegeln zum Teil Überschätzungen der durch die PISA-Ergebnisse nahegelegten Unterschiede zwischen den Schülergruppen mit den betreffenden askriptiven Merkmalen wider. Daher kann vermutet werden, dass ein Großteil der Lehrkräfte aus der untersuchten Stichprobe bei den Einschätzungen auf unzutreffende ethnische, soziale und geschlechtsspezifische Stereotype zurückgegriffen hat. Dies wiederum impliziert, dass die im Rahmen der Erhebungen gemessenen Lehrerurteile zu großen Teilen die Folge kategorisierender Wahrnehmungsprozesse gewesen sein könnten. Eine direkte Untersuchung dieser Vermutung ist mit den zugrundeliegenden Daten nicht möglich.

Teilweise wurden aber auch Leistungserwartungen belegt, die auf Gruppenebene zutreffend sind. Für dieses Muster lassen sich verschiedene Erklärungen anführen. So wäre es denkbar, dass bestimmte Schülergruppen tendenziell eher anhand kategorisierender Informationsverarbeitungsstile eingeschätzt wurden als andere Gruppen, die eher im Rahmen von merkmalsbasierten Prozessen wahrgenommen wurden. In diesem Zusammenhang sei darauf verwiesen, dass sich eigenschafts- und kategorienbasierte Verarbeitungsstile nicht gegenseitig ausschließen, sondern die Endpunkte eines Kontinuums an Verarbeitungsstilen bilden. Somit könnten zutreffende Einschätzungen zukünftiger Mathematikleistungen von Mädchen und von türkischstämmigen Kindern die Folge von Verarbeitungsprozessen gewesen sein, die zwar nicht ausschließlich, dafür aber zu einem höheren Anteil eigenschaftsbasiert waren als die den Lehrererwartungen in Bezug auf das Fach Deutsch zugrunde liegenden Verarbeitungsprozesse. Gleichzeitig kommt als Erklärung der gefundenen Muster in Betracht, dass bestimmte Stereotype die Merkmale der betreffenden Schülergruppen korrekt vorhersagen und deshalb zutreffend sind, während dies für andere Stereotype, die eher unzutreffend sind, nicht der Fall ist.

Vor dem Hintergrund der Zielsetzung individuelle und gruppenbezogene Verzerrungen von Lehrererwartungen (sowie mögliche Auswirkungen in Form einer selbsterfüllenden Prophezeiung) zu vermeiden, ist eine weiterführende Erforschung

der Frage notwendig, ob und unter welchen Bedingungen Einschätzungen von Lehrkräften auf der Grundlage kategorisierender Informationsverarbeitungsprozesse getroffen werden und wann es zu einer Verlagerung der Wahrnehmung hin zu eher eigenschaftsbasierten Stilen kommt. Der Abbau von unzutreffenden Stereotypen bei Lehrkräften scheint unabhängig davon in jedem Falle eine mögliche Strategie zu sein, um systematische Verzerrungen in den Erwartungen von Lehrkräften zu reduzieren.

Kritisch sei darauf hingewiesen, dass den empirischen Analysen ein recht enges Konzept zutreffender Lehrererwartungen zugrunde liegt. Dabei wurden Unterschiede in den Lehrererwartungen, die sich nicht auf gemessene Leistungen und Fähigkeiten sowie die Motivation der Kinder zurückführen ließen, im Sinne einer Residualinterpretation als Verzerrungen gedeutet. Dieses Vorgehen setzt die nur schwer überprüfbare Annahme voraus, dass sich über die verwendeten Messinstrumente alle für die Einschätzungen von Lehrkräften relevanten Leistungsaspekte abbilden lassen. Denkbar ist jedoch, dass sich die betrachteten Schülergruppen zusätzlich auch in anderen, nicht gemessenen Merkmalen voneinander unterscheiden und dass Lehrkräfte diese Informationen für eine aus ihrer Sicht zutreffende Einschätzung der zukünftigen Schülerleistungen heranziehen. Beispielsweise könnten Unterschiede im Lern- und Leistungsverhalten der Kinder im Unterricht, welche im Rahmen dieses Beitrags nicht berücksichtigt werden konnten, für eine Erklärung systematischer Verzerrungen in den Lehrererwartungen von Bedeutung sein. So könnten die gefundenen positiv verzerrten Erwartungen an die Leistungen von Mädchen im Fach Deutsch auf nicht kontrollierte Merkmale wie beispielsweise die Gewissenhaftigkeit zurückzuführen sein. Ebenso könnte aus dem Blickwinkel einer etwas weiter gefassten Konzeption eingewendet werden, dass an der sozioökonomischen Lage ausgerichtete Lehrererwartungen eigentlich nicht verzerrt sind, da die Lehrkräfte hohe elterliche Unterstützungsleistungen für Kinder aus bessergestellten Familien antizipieren.⁶ Ob Lehrkräfte ein breiteres Feld an leistungsrelevanten Merkmalen heranziehen, sollte anhand weiterführender Forschung erörtert werden.

Weiterhin ist zu berücksichtigen, dass sich die Schulen und Lehrkräfte freiwillig für die Teilnahme an der Studie entschieden haben (bei einer gleichzeitig relativ hohen Anzahl an Absagen). Dies könnte zu einer Auswahl an Lehrkräften geführt haben, die im Vergleich zum Durchschnitt besonders engagiert ist. Wenn allerdings davon ausgegangen wird, dass sich diese Lehrkräfte vor dem Hintergrund ihres Engagements um besonders angemessene Einschätzungen bemühen, so sollten die gefundenen Verzerrungen tendenziell eine Unterschätzung der Effekte in der Grundgesamtheit darstellen. Die zu erwartende soziale Erwünschtheit im Rahmen von Befragungen legt eine ähnliche Vermutung nahe.

Eine weitere Einschränkung betrifft die zusammengefasste Betrachtung verschiedener Zuwanderergenerationen. Da Merkmale wie geringere sprachliche Kompetenzen in der ersten Generation in stärkerem Maße anzutreffen sein dürften als in der zweiten oder dritten, sollte die Aktivierung von Stereotypen in der Generationenfolge variieren. Entsprechend wäre zu erwarten, dass die festgestellten

⁶Allerdings sollten die in den Analysen berücksichtigten Fähigkeitsindikatoren elterliche Unterstützungsleistungen teilweise abbilden, weil diese die schulrelevanten Fähigkeiten bereits in der Zeit vor den Leistungsmessungen maßgeblich beeinflusst haben sollten.

Verzerrungen im Vergleich zur zweiten und dritten Generation in einem stärkeren Maße in der ersten Generation anzutreffen sind. Der Ausschluss der ersten Generation aus den Analysen sollte demnach ebenfalls zu einer Unterschätzung der ethnischen Verzerrungen in den Lehrererwartungen führen.

Außerdem sollte nicht unerwähnt bleiben, dass die Stichprobe im Vergleich zu repräsentativen Stichproben im Hinblick auf ihre sozioökonomische Zusammensetzung tendenziell positiv selektiv ist. Das zeigen die durchschnittlich hohen HISEI-Werte (Tab. 1). Dass sich Verzerrungen in Lehrererwartungen nach der sozialen Herkunft der Kinder trotz dieser Tatsache zeigen, könnte bedeuten, dass Verzerrungen in den Lehrererwartungen nach der sozialen Herkunft ebenfalls unterschätzt wurden. Da die vermutete Selektivität der Stichprobe auch bedeuten würde, dass die türkisch-stämmige Zuwanderergruppe positiv selektiv ist, könnte ähnliches auch auf das Ausmaß an Verzerrungen gegenüber dieser Gruppe zutreffen.

Die vorliegende Studie liefert insgesamt Hinweise darauf, dass Leistungserwartungen von Lehrkräften zum Teil systematisch nach dem sozialen und ethnischen Hintergrund sowie dem Geschlecht der Kinder verzerrt sind. Damit wäre eine zentrale Voraussetzung für die Entstehung selbsterfüllender Prophezeiungen im schulischen Kontext gegeben, die zu einer Verfestigung oder gar Verstärkung sozialer, ethnischer und geschlechtsspezifischer Bildungsungleichheiten beitragen könnten. Ein derartiger Prozess würde entstehen, wenn sich die Interaktionen im Klassenzimmer durch die verzerrten Erwartungen der Lehrenden so verändern, dass die Leistungsentwicklung der Schüler diesen Erwartungen zu späteren Zeitpunkten entspricht. Weiterführende Forschungsarbeiten, die sich mit den Auswirkungen von Lehrererwartungen auf das Verhalten von Lehrkräften und die Kompetenzentwicklung von Schülern beschäftigen, sind deshalb dringend erforderlich.

Literatur

- Alvidrez, Jennifer, und Rhona S. Weinstein. 1999. Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology* 91:731–746.
- Arnold, Eva. 2001. Sozialpsychologische Beiträge zur Diskussion um die Zukunft der Schulen in Deutschland. *Gruppendynamik und Organisationsberatung* 32:71–82.
- Arnold, Karl-Heinz, Wilfried Bos, Peggy Richert und Tobias C. Stubbe. 2007. Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In *IGLU 2006: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Hrsg. Wilfried Bos, Sabine Hornberg, Karl-Heinz Arnold, Gabriele Faust, Lilian Fried, Eva-Maria Lankes, Knut Schwippert und Renate Valtin, 271–297. Münster: Waxmann.
- Arrow, Kenneth. 1973. The theory of discrimination. In *Discrimination in labor markets*, Hrsg. Orley Ashenfelter und Albert Rees, 3–33. Princeton: Princeton University Press.
- Baron, Reuben M., David Y. H. Tom und Harris M. Cooper. 1985. Social class, race and teacher expectations. In *Teacher expectancies*, Hrsg. Jerome B. Dusek, 251–269. Hillsdale: Lawrence Erlbaum.
- Bäuerlein, Kerstin, Andrea Beinicke, Nicole Berger, Gabriele Faust, Melanie Jost und Wolfgang Schneider. 2012. *FIPS – Fähigkeitsindikatoren Primarschule*. Göttingen: Hogrefe.
- Baumert, Jürgen, Petra Stanat und Rainer Watermann. Hrsg. 2006. *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, Dominik. 2013. The impact of teachers' expectations on students' educational opportunities in the life course: An empirical test of a subjective expected utility explanation. *Rationality and Society* 25:422–469.

- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther. Hrsg. 2004. *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Brewer, Marilynn B. 1988. *A dual process model of impression formation*. Hillsdale: Lawrence Erlbaum Associates, Inc.
- Diehl, Claudia, und Patrick Fick. 2016. Ethnische Diskriminierung im deutschen Bildungssystem. In *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*, Hrsg. Claudia Diehl, Christian Hunkler und Cornelia Kristen, 243–286. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Doherty, Jim, und Michael Conolly. 1985. How accurately can primary school teachers predict the scores of their pupils in standardised tests of attainment? A study of some non-cognitive factors that influence specific judgements. *Educational Studies* 11:41–60.
- Dumke, Dieter. 1977. Die Auswirkungen von Lehrererwartungen auf Intelligenz und Schulleistungen. *Psychologie in Erziehung und Unterricht* 24:93–108.
- Dumont, Hanna, Kai Maaz, Marko Neumann und Michael Becker. 2014. Soziale Ungleichheiten beim Übergang von der Grundschule in die Sekundarstufe I: Theorie, Forschungsstand, Interventions- und Fördermöglichkeiten. *Zeitschrift für Erziehungswissenschaft* 17:141–165.
- Dusek, Jerome B., und Gail Joseph. 1983. The bases of teacher expectancies. *Journal of Educational Psychology* 75:327–346.
- Eccles, Jacquelynne S., und Phyllis Blumenfeld. 1985. Classroom experiences and student gender: Are there differences and do they matter. In *Gender influences in classroom interaction*, Hrsg. Louise C. Wilkinson und Cora B. Marrett, 79–114. Orlando: Academic Press.
- Ehrenberg, Ronald G., Daniel D. Goldhaber und Dominic J. Brewer. 1995. Do teachers' race, gender, and ethnicity matter? Evidence from NELS88. *Industrial and Labor Relations Review* 48:547–561.
- England, Paula. 1992. *Comparable worth: Theories and evidence*. New York: Transaction Publishers.
- Esser, Hartmut. 2006. *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt a. M.: Campus.
- Fiske, Susan T. 1998. Stereotyping, prejudice, and discrimination. In *The handbook of social psychology, Vols. 1 and 2* (4. Aufl.), Hrsg. Gilbert D. T. Gilbert, Susan T. Fiske und Gardner Lindzey. New York: McGraw-Hill.
- Fiske, Susan T., Monica Lin und Steven L. Neuberg. 1999. The continuum model. In *Dual-process theories in social psychology*, Hrsg. Shelly Chaiken und Yaacov Trope, 321–254. New York: The Guilford Press.
- Florack, Arnd, Martin Scarabis und Herbert Bless. 2001. When do associations matter? The use of automatic associations toward ethnic groups in person judgments. *Journal of Experimental Social Psychology* 37:518–524.
- Friedrich, Alena, Barbara Flunger, Benjamin Nagengast, Kathrin Jonkmann und Ulrich Trautwein. 2015. Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology* 41:1–12.
- Ganzeboom, Harry B. G. 2010. A New International Socio-Economic Index (ISEI) of Occupational Status for the International Standard Classification of Occupation 2008 (ISCO-08) Constructed with data from the ISSP 2002–2007. Annual Conference of International Social Survey Programme. Lissabon.
- Ganzeboom, Harry B. G., Paul M. De Graaf und Donald J. Treiman. 1992. A standard international socio-economic index of occupational status. *Social Science Research* 21:1–56.
- Glock, Sabine, und Sabine Krolak-Schwerdt. 2013. Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Social Psychology of Education* 16:111–127.
- Gresch, Cornelia. 2012. *Der Übergang in die Sekundarstufe I*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Halpern, Diane F. 2013. *Sex differences in cognitive abilities*. New York: Psychology Press.
- Hannover, Bettina. 2004. Gender revisited: Konsequenzen aus PISA für die Geschlechterforschung. *Zeitschrift für Erziehungswissenschaft* 7:81–99.
- Hartley, David. 1982. Ethnicity or sex: Teachers' definitions of ability and reading comprehension in an EPA primary school. *Research in Education* 28:9–24.
- Hinnant, J. B., Marion O'Brien und Sharon R. Ghazarian. 2009. The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology* 101:662–670.

- Hirschauer, Maria, und Harry Kullmann. 2010. Lehrerprofessionalität im Zeichen von Heterogenität. Stereotype bei Lehrkräften als kollegial zu bearbeitende Herausforderung. In *Ethnizität, Geschlecht, Familie und Schule*, Hrsg. Maria Hirschauer und Harry Kullmann, 351–373. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hoge, Robert D., und Robert Butcher. 1984. Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology* 76:777–781.
- Hoge, Robert D., und Theodore Coladarsi. 1989. Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research* 59:297–313.
- Jussim, Lee. 1989. Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology* 57:469–480.
- Jussim, Lee, und Jacquelynne S. Eccles. 1992. Teacher expectations: II. Construction and reflection of student achievement. *Journal of Personality and Social Psychology* 63:947–961.
- Jussim, Lee, und Jacquelynne S. Eccles. 1995. Are teacher expectations biased by students' gender, social class, or ethnicity? In *Stereotype accuracy. Toward appreciating group differences*, Hrsg. Yueh-Ting Lee, Lee J. Jussim und Clark R. McCauley, 245–271. Washington, D.C.: American Psychological Association.
- Jussim, Lee, Jacquelynne S. Eccles und Stephanie Madon. 1996. Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology* 28:281–388.
- Jussim, Lee, und Kent D. Harber. 2005. Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review* 9: 131–155.
- Jussim, Lee, Stacy L. Robustelli und Thomas R. Cain. 2009. Teacher expectations and self-fulfilling prophecies. In *Handbook of motivation at school*, Hrsg. Kathryn R. Wentzel und Allan Wigfield, 349–380. New York: Routledge.
- Kristen, Cornelia. 2008. Schulische Leistungen von Kindern aus türkischen Familien am Ende der Grundschulzeit. Befunde aus der IGLU-Studie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie Sonderheft* 48:230–251.
- Kristen, Cornelia, und Nadia Granato. 2007. The educational attainment of the second generation in Germany. Social origins and ethnic inequality. *Ethnicities* 7:343–366.
- Lüdemann, Elke, und Guido Schwerdt. 2013. Migration background and educational tracking. *Journal of Population Economics* 26:455–481.
- Ludwig, Peter H., und Heidrun Ludwig. Hrsg. 2007. *Erwartungen in himmelblau und rosarot. Effekte, Determinanten und Konsequenzen von Geschlechterdifferenzen in der Schule*. Weinheim: Juventa Verlag.
- Madon, Stephanie, Lee J. Jussim und Jacquelynne S. Eccles. 1997. In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology* 72:791–809.
- Madon, Stephanie, Lee J. Jussim, Shelley Keiper, Jacquelynne S. Eccles, Alison Smith und Polly Palumbo. 1998. The accuracy and power of sex, social class, and ethnic stereotypes: A naturalistic study in person perception. *Personality and Social Psychology Bulletin* 24:1304–1318.
- McKown, Clark, und Rhona S. Weinstein. 2002. Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology* 32:159–184.
- McKown, Clark, und Rhona S. Weinstein. 2008. Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology* 46:235–261.
- Müller, Andrea G., und Petra Stanat. 2006. Schulischer Erfolg von Schülerinnen und Schülern mit Migrationshintergrund: Analysen zur Situation von Zuwanderern aus der ehemaligen Sowjetunion und aus der Türkei. In *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit*, Hrsg. Jürgen Baumert, Petra Stanat und Rainer Watermann, 221–255. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Muthén, Linda K., und Bengt O. Muthén. 2012. *Mplus user's guide*. 7. Aufl. Los Angeles: Muthén & Muthén.
- Olczyk, Melanie, Gisela Will und Cornelia Kristen. 2014. *Personen mit Zuwanderungshintergrund im NEPS: Zur Bestimmung von Generationenstatus und Herkunftsgruppe (NEPS Working Paper No. 41b)*. Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.
- Parsons, Jacquelynne S., Caroline M. Kaczala und Judith L. Meece. 1982. Socialization of achievement attitudes and beliefs: Classroom influences. *Child Development* 53:322–339.

- Phelps, Edmund S. 1972. The statistical theory of racism and sexism. *The American Economic Review* 62:659–661.
- Prenzel, Manfred, und Jürgen Baumert. 2004. PISA 2003: Der Bildungsstandard der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. In *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*, Hrsg. PISA-Konsortium Deutschland. Münster: Waxmann.
- Rauer, Wulf, und Karl D. Schuck. 2004. *Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen (FEES I–2)*. Göttingen: Beltz Verlag.
- Ready, Douglas D., und Elizabeth M. Chu. 2015. Sociodemographic inequality in early literacy development: The role of teacher perceptual accuracy. *Early Education and Development* 26:1–18.
- Ready, Douglas D., und David L. Wright. 2011. Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities. The role of child background and classroom context. *American Educational Research Journal* 48:335–360.
- Rosenthal, Robert, und Lenore Jacobson. 1968. Pygmalion in the classroom. *The Urban Review* 3:16–20.
- Ross, Sandra I., und Jeffrey M. Jackson. 1991. Teachers' expectations for black males' and black females' academic achievement. *Personality and Social Psychology Bulletin* 17:78–82.
- Schafer, Joseph L., und John W. Graham. 2002. Missing data: Our View of the state of the art. *Psychological Methods* 7:147–177.
- Schofield, Janet W. 2006. *Migration background, minority-group membership and academic achievement. Research evidence from social, educational, and developmental psychology*. Berlin: AKI Research Review 5.
- Segeritz, Michael, Oliver Walter und Petra Stanat. 2010. Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland: Evidenz für segmentierte Assimilation? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62:113–138.
- Sherman, Jeffrey W., Steven J. Stroessner, Frederica R. Conrey und Omar A. Azam. 2005. Prejudice and stereotype maintenance processes: Attention, attribution, and individuation. *Journal of Personality and Social Psychology* 89:607–622.
- Smith, Eliot R., und Jamie DeCoster. 2000. Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review* 4:108–131.
- Snijders, Tom, und Roel Bosker. 1999. *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: Sage.
- Snyder, Mark, und Arthur A. Stukas. 1999. Interpersonal processes: The interplay of cognitive, motivational, and behavioral activities in social interaction. *Annual Review of Psychology* 50:273–303.
- Spitz, Herman H. 1999. Beleaguered pygmalion: A history of the controversy over claims that teacher expectancy raises intelligence. *Intelligence* 27:199–234.
- Sprietsma, Maresa. 2013. Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics* 45: 523–538.
- Stahl, N. 2007. Schülerwahrnehmung und -beurteilung durch Lehrkräfte. In *Kompetenzaufbau und Laufbahnen im Schulsystem. Ergebnisse einer Längsschnittuntersuchung an Grundschulen*, Hrsg. Hartmut Ditton, 171–198. Münster: Waxmann.
- Stanat, Petra, und Susanne Bergann. 2009. Geschlechtsbezogene Disparitäten in der Bildung. In *Handbuch Bildungsforschung*, Hrsg. Rudolf Tippelt und Bernhard Schmidt, 513–527. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stanat, Petra, und Gayle Christensen. 2006. *Schulerfolg von Jugendlichen mit Migrationshintergrund im internationalen Vergleich: Eine Analyse von Voraussetzungen und Erträgen schulischen Lernens im Rahmen von PISA 2003*. Bonn: Bundesministerium für Forschung und Bildung (BMBF).
- Stubbe, Tobias C., und Wilfried Bos. 2008. Schullaufbahneempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik* 22:49–63.
- Tenenbaum, Harriet R., und Martin D. Ruck. 2007. Are teachers' expectations different for racial minority than for European American students? A Meta-analysis. *Journal of Educational Psychology* 99:253–273.
- Ulich, Klaus. 2001. Die Lehrer/innen-Schüler/innen-Interaktion. In *Einführung in die Sozialpsychologie der Schule*, Hrsg. Klaus Ulich, 76–115. Weinheim: Belt Verlag.
- Weiß, Rudolf H., und Jürgen Osterland. 1997. *Grundintelligenztest Skala 1 – CFT1*. Göttingen: Hogrefe.

Georg Lorenz, 1982, Berliner Institut für empirische Integrations- und Migrationsforschung (BIM), Humboldt-Universität zu Berlin. Forschungsgebiete: Migrations- und Integrationsprozesse; Diskriminierung; Bildungssoziologie.

Sarah Gentrup, 1987, Institut für Erziehungswissenschaften an der Humboldt-Universität zu Berlin in Kooperation mit dem Institut zur Qualitätsentwicklung im Bildungswesen (IQB). Forschungsgebiete: Ethnische, soziale und geschlechtsbezogene Disparitäten im Bildungserfolg; Lehrkraft-Kind-Interaktionen; Urteilsprozesse.

Cornelia Kristen, 1972, Prof. Dr., rer. soc., Lehrstuhl für Soziologie, insbesondere Sozialstrukturanalyse an der Universität Otto-Friedrich-Universität Bamberg. Forschungsgebiete: Migrations- und Integrationsprozesse; Bildungssoziologie; Soziale Ungleichheit. Ausgewählte Veröffentlichungen: Gender inequalities in the education of the second generation in western countries. *Sociology of Education* 87, 2014 (mit F. Fleischmann); Primary school choice and ethnic school segregation in German elementary schools. *European Sociological Review* 24, 2008.

Petra Stanat, 1964, Prof. Dr., Direktorin des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) an der Humboldt-Universität zu Berlin. Forschungsgebiete: Bedingungen und Förderung des Bildungserfolgs von Kindern und Jugendlichen mit Migrationshintergrund; Ethnische, soziale und geschlechtsbezogene Disparitäten im schulischen Erfolg; Zweitsprachförderung und Lesekompetenz; Schulleistungsvergleiche; Sicherung von Mindeststandards. Veröffentlichungen: Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik: Ergebnisse des IQB-Ländervergleichs 2011, Münster 2012 (als Hrsg. mit H. A. Pant, K. Böhme und D. Richter); Improving second language skills of immigrant students: A field trial study evaluating the effects of a summer learning program. *Learning and Instruction* 22, 2012 (mit M. Becker, J. Baumert, O. Lüdtke und A. G. Eckhardt).

Irena Kogan, 1973, Prof. Dr. soc., Professur für Soziologie, Gesellschaftsvergleich an der Universität Mannheim. Forschungsgebiete: Integration von Zuwanderern und ethnische Schichtung; Soziale Ungleichheit; Übergang vom Bildungssystem in den Arbeitsmarkt; Vergleichende Sozialforschung. Aktuellste Veröffentlichungen: Migrant Networks and Labor Market Integration of Immigrants from the Former Soviet Union in Germany. *Social Forces* 92, 2014 (mit F. Kalter); Relative worth of a bachelor degree: Patterns of labour market integration among drop-outs and graduates in sequential and integrated tertiary education systems. *Acta Sociologica* 57, 2014 (mit T. Matković); Immigrants' initial steps in Germany and their later economic success. *Advances in Life Course Research* 18 (mit M. Weissmann).

Teilstudie 2: Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen

Dieser Beitrag ist in der *Zeitschrift für Erziehungswissenschaft* erschienen. Die Referenz lautet:

Gentrup, S., Rjosk, C., Stanat, P. & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen [Teachers' perceptions of students' motivation and learning behaviour and their role in biased teacher achievement expectations]. *Zeitschrift für Erziehungswissenschaft*, 21, 867–891. Doi: 10.1007/s11618-018-0806-2

The final publication is available at Springer Nature via

<http://dx.doi.org/10.1007/s11618-018-0806-2>

Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen

Sarah Gentrup  · Camilla Rjosk · Petra Stanat · Georg Lorenz

Online publiziert: 26. Januar 2018
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

Zusammenfassung Verschiedene Studien haben gezeigt, dass Leistungserwartungen von Lehrkräften teilweise in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund sowie vom Geschlecht der Schülerinnen und Schüler stereotyp verzerrt sein können. Im vorliegenden Beitrag wird die Annahme untersucht, dass diese systematischen Verzerrungen in Leistungserwartungen im sprachlichen und mathematischen Bereich teilweise auf die Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens der Schülerinnen und Schüler zurückgeführt werden können. Die Ergebnisse, die auf einer Stichprobe von $N=901$ Kindern aus $N=66$ ersten Grund-

Das diesem Beitrag zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 01JC1117A (Otto-Friedrich-Universität Bamberg), 01JC1117B (Universität Mannheim) und 01JC1117C (Humboldt-Universität zu Berlin) gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s11618-018-0806-2>) enthalten.

S. Gentrup (✉)
Institut für Erziehungswissenschaften, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: sarah.gentrup@hu-berlin.de

Dr. C. Rjosk · Prof. Dr. P. Stanat · Dr. G. Lorenz
Institut zur Qualitätsentwicklung im Bildungswesen (IQB) - Wissenschaftliche Einrichtung der Länder an der Humboldt-Universität zu Berlin e.V., Unter den Linden 6, 10099 Berlin, Deutschland

Dr. C. Rjosk
E-Mail: camilla.rjosk@iqb.hu-berlin.de

Prof. Dr. P. Stanat
E-Mail: iqboffice@iqb.hu-berlin.de

Dr. G. Lorenz
E-Mail: georg.lorenz@iqb.hu-berlin.de

schulklassen mit 69 Klassen- und Fachlehrkräften basieren, unterstützen dies für soziale und geschlechtsbezogene Verzerrungen. Verzerrungen in Abhängigkeit vom ethnischen Hintergrund der Kinder scheinen hingegen unabhängig von der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens der Schülerinnen und Schüler aufzutreten. Rückschlüsse auf stereotype Annahmen und deren Bedeutung für unterrichtliches Handeln sowie für Bildungsgleichheit werden diskutiert.

Schlüsselwörter Leistungsmotivation · Arbeitsverhalten · Lehrererwartungen · Lehrerurteile · Stereotype

Teachers' perceptions of students' motivation and learning behaviour and their role in biased teacher achievement expectations

Abstract Several studies have shown that teachers' achievement expectations for their students may be biased to some extent by students' ethnic or social background as well as gender. The present study investigates whether these biases in teachers' expectations of students' performance in early language and mathematics are partly explained by differences in teachers' perceptions of students' motivation and learning behaviour. The analyses are based on data from a sample of $N=901$ first-grade students from $N=66$ classrooms with 69 teachers. The results support the hypothesis for biases associated with social background and gender. Ethnic bias, however, occurred independently of differences in perceived motivation and learning behaviour. We discuss the relevance of stereotypical assumptions for teaching as well as educational disparities.

Keywords Achievement motivation · Learning behaviour · Stereotypes · Teacher expectations · Teacher judgements

1 Einleitung

Die Existenz ethnischer, sozialer und geschlechtsbezogener Bildungsungleichheiten ist empirisch gut belegt (vgl. z.B. Prenzel et al. 2013). Zur Erklärung dieser Disparitäten werden in der empirischen Bildungsforschung verschiedene Faktoren diskutiert, wie etwa Struktur- und Prozessmerkmale der Familien (Baumert et al. 2003), Sprachkenntnisse der Schülerinnen und Schüler (Heinze et al. 2007) sowie verschiedene Arten von Diskriminierung (Diehl und Fick 2016). Eine weitere Annahme bezieht sich auf Leistungserwartungen von Lehrkräften, die für Bildungsgleichheit bedeutsam sein können, da sie eine wichtige Grundlage für pädagogisches Handeln im Unterricht bilden (Schrader und Helmke 2001).

Leistungserwartungen von Lehrkräften basieren nur zum Teil auf tatsächlichen Leistungen der Lernenden (z.B. Südkamp et al. 2012) und scheinen auch unter Kontrolle der Leistungen in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund oder Geschlecht der Heranwachsenden unterschiedlich auszufallen (z.B. Jussim und Harber 2005; Ready und Wright 2011; Lorenz et al. 2016). Obwohl solche systematischen Abweichungen in Leistungserwartungen von Lehrkräften, die

als Verzerrungen bezeichnet werden können, sowohl im internationalen als auch im deutschen Kontext wiederholt belegt wurden, ist über ihre Ursachen wenig bekannt. Angenommen werden kann, dass Lehrkräfte in der Wahrnehmung einzelner Kinder unbewusst und automatisch auf Stereotype, also auf verallgemeinernde Vorstellungen über Merkmale, Eigenschaften und Verhaltensweisen von Schülergruppen zurückgreifen (vgl. Fiske und Neuberg 1990; Lorenz 2017). Stereotype von Lehrkräften können sich dabei auf verschiedene Inhalte beziehen. Denkbar wären beispielsweise verallgemeinernde Vorstellungen über verschiedene Fähigkeiten von Lernenden oder über die häusliche Unterstützung, die bestimmte Gruppen von Schülerinnen und Schülern erfahren. Studien zu Übergangsempfehlungen zeigen allerdings, dass Lehrkräfte neben schulischen Leistungen primär der Motivation und dem Arbeitsverhalten von Lernenden hohe Relevanz für zukünftige Leistungen beimessen (z. B. Böhmer et al. 2015). Als mögliche Quelle für Verzerrungen von Leistungserwartungen kommen daher insbesondere stereotypenbasierte Annahmen über die Motivation und das Arbeitsverhalten der Kinder in Betracht. Der vorliegende Beitrag untersucht deshalb, ob ethnische, soziale und geschlechtsbezogene Verzerrungen in Leistungserwartungen auf Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens ihrer Schülerinnen und Schüler zurückgeführt werden können. Hinausgehend über viele der bisherigen Studien werden dabei in der vorliegenden Arbeit Selbsteinschätzungen der Kinder zu ihrer Motivation und ihrem Arbeitsverhalten berücksichtigt und als Vergleichsmaß für die Lehrkräfteeinschätzungen herangezogen. Anhand dieses Vorgehens lässt sich bestimmen, ob die Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens primär die Selbsteinschätzungen der Lernenden oder eher stereotypenbasierte Annahmen widerspiegeln.

2 Theorie und Forschungsstand

2.1 Leistungserwartungen von Lehrkräften und deren Verzerrungen

Bei Leistungserwartungen von Lehrkräften handelt es sich um Prognosen der zukünftigen Leistungen von Schülerinnen und Schülern. Auch wenn sie sich konzeptionell von Leistungseinschätzungen aktueller Fähigkeiten abgrenzen lassen, werden in der Forschung häufig Einschätzungen aktueller Leistungen als Indikatoren für Leistungserwartungen herangezogen und beide Begriffe synonym gebraucht. Um die Lesbarkeit zu erhöhen, wird im Folgenden zusammenfassend von *Leistungserwartungen* gesprochen. Ein weiteres verwandtes Konzept ist die diagnostische Kompetenz von Lehrkräften. Hiermit ist die übergeordnete Fähigkeit gemeint, zu genauen Urteilen über Merkmale von Lernenden, wie z.B. Leistungen, zu gelangen (vgl. Schrader 2010). Um den Forschungsstand umfassend darzustellen, werden im Folgenden einschlägige Ergebnisse aller drei Konzeptionen berücksichtigt. Für die Operationalisierung der Leistungserwartungen werden in der vorliegenden Arbeit jedoch ausschließlich Prognosen zukünftiger Leistungen der Schülerinnen und Schüler herangezogen.

Zahlreiche Studien zeigen, dass Leistungserwartungen von Lehrkräften zu einem hohen Anteil die tatsächlichen Leistungen von Heranwachsenden widerspiegeln

(Jussim et al. 1996). Laut Metaanalysen beträgt die geteilte Varianz von Leistungserwartungen und standardisiert erfassten Leistungen etwa 40 % (Hoge und Coladarci 1989; Südkamp et al. 2012). Dies bedeutet jedoch auch, dass ein substanzieller Varianzanteil der Erwartungen nicht auf die Leistungen zurückgeführt werden kann.

Verschiedene Studien verdeutlichen zudem, dass die beobachteten Über- und Unterschätzungen nicht rein zufällig sind. Vielmehr traten über Leistungsunterschiede hinaus systematische Unterschiede in den Leistungserwartungen von Lehrkräften für verschiedene Schülergruppen auf (vgl. auch Lorenz et al. 2016). Derartige Unterschiede werden im Folgenden als *Verzerrungen* bezeichnet.

Internationale Studien weisen darauf hin, dass Leistungserwartungen von Lehrkräften an Kinder aus zugewanderten Familien (z. B. Jussim und Harber 2005; McKown und Weinstein 2008) und an Kinder aus sozial benachteiligten Familien negativ verzerrt sind (z. B. Jussim et al. 1996; Ready und Wright 2011). Die Befunde zu geschlechtsbezogenen Verzerrungen sind hingegen uneinheitlich und deuten auf bereichsspezifische Effekte hin. So fanden einige Studien eine Überschätzung von Jungen im mathematischen (Jussim und Eccles 1992) und von Mädchen im sprachlichen Bereich (Hinnant et al. 2009; Ready und Wright 2011). Andere Untersuchungen identifizierten keine Verzerrungen in Abhängigkeit vom Geschlecht (McKown und Weinstein 2008; van den Bergh et al. 2010).

In Deutschland wurden bisher kaum Studien zu Verzerrungen in Leistungserwartungen durchgeführt. In Untersuchungen zur diagnostischen Kompetenz von Lehrkräften zeigten sich nach Kontrolle standardisierter Leistungsmaße keine Unterschiede in den Leistungsurteilen für Kinder mit und ohne Zuwanderungshintergrund (Schrader und Helmke 1990; Stahl 2007; Hachfeld et al. 2010; Rjosk et al. 2011). Diese Analysen differenzierten jedoch nicht nach ethnischen Herkunftsgruppen, für die unterschiedliche Stereotype und darauf basierende Annahmen bestehen könnten.

Zu sozialen und geschlechtsbezogenen Verzerrungen in Leistungsurteilen sind die Befunde im deutschsprachigen Raum uneinheitlich. In einigen Studien wurden keine sozialen (Hachfeld et al. 2010; Karing et al. 2011) oder geschlechtsbezogenen Verzerrungen (Schrader und Helmke 1990; Karing et al. 2011) identifiziert. Stahl (2007) berichtete hingegen Überschätzungen im Bereich Lesen für Kinder aus sozial bessergestellten Familien und für Mädchen.

Aufgrund der uneinheitlichen und insgesamt unzureichenden Befundlage, sind Lorenz et al. (2016) anhand von Daten der in Deutschland durchgeführten KuL-Studie der Frage nachgegangen, inwieweit Leistungserwartungen von Lehrkräften in Abhängigkeit vom ethnischen und sozialen Hintergrund sowie vom Geschlecht der Lernenden verzerrt sind. Im Fach Deutsch erwarteten die Lehrkräfte bei gleichen Testleistungen der Erstklässlerinnen und Erstklässler eine weniger günstige Leistungsentwicklung für türkischstämmige Kinder im Vergleich zu Kindern ohne Zuwanderungshintergrund, für Kinder aus sozial schwächeren Familien im Vergleich zu Kindern aus sozial bessergestellten Familien sowie für Jungen im Vergleich zu Mädchen. Im Fach Mathematik fielen die Leistungserwartungen bei gleichen tatsächlichen Leistungen höher aus für Kinder mit osteuropäischem Hintergrund als für Kinder ohne Zuwanderungshintergrund und höher für Kinder aus sozial bessergestellten Familien als für Kinder aus sozial benachteiligten Familien. Geschlechtsbezogene Verzerrungen waren im Fach Mathematik nicht zu verzeichnen.

Zusammenfassend scheinen Leistungserwartungen von Lehrkräften nicht nur die tatsächlichen Leistungen ihrer Schülerinnen und Schüler widerzuspiegeln. Außerdem wird deutlich, dass Leistungserwartungen in Abhängigkeit von Hintergrundmerkmalen der Lernenden verzerrt sein können. Offen ist allerdings, worauf diese Verzerrungen zurückzuführen sind und ob ihnen stereotype Annahmen über weitere lernrelevante Merkmale der Lernenden, wie etwa deren Motivation und Arbeitsverhalten, zugrunde liegen.

2.2 Stereotype Wahrnehmungsprozesse

Als theoretischer Rahmen zur Erklärung systematischer Verzerrungen in Leistungserwartungen eignet sich das sozialpsychologische Kontinuum-Modell (Fiske und Neuberg 1990). Diesem Modell zufolge bewegt sich die Personenwahrnehmung zwischen zwei Polen – der individualisierenden und der kategorisierenden Verarbeitung. Bei letzterer werden Personen anhand eines einfach zu identifizierenden Merkmals einer sozialen Kategorie zugeordnet (Ganter 1997), wobei die Merkmale Alter, Geschlecht, ethnische Herkunft sowie sozialer Status besonders salient sind (Fiske und Neuberg 1990). Es wird angenommen, dass die Zuordnung einer Person zu einer Gruppe kognitive Stereotype aktiviert, also verallgemeinernde Vorstellungen über Merkmale, Eigenschaften und Verhaltensweisen der Gruppenmitglieder (Ganter 1997). Im Wahrnehmungsprozess findet im Allgemeinen zuerst eine kategorisierende Verarbeitung statt. Eine Verschiebung hin zur individualisierenden Verarbeitung, bei der individuelle Eigenschaften und Merkmale der jeweiligen Person berücksichtigt werden, erfolgt erst unter bestimmten Voraussetzungen wie z. B. dem Vorliegen stereotypen-inkonsistenter Informationen (Fiske und Neuberg 1990; Ganter 1997). Ohne eine derartige Verschiebung wird die weitere Wahrnehmung durch Stereotype gelenkt.

Auch in schulischen Situationen können stereotype Wahrnehmungsprozesse auftreten, etwa wenn Einschätzungen von Lehrkräften durch stereotype Merkmalsannahmen für bestimmte Schülergruppen gefärbt sind (Hirschauer und Kullmann 2010). Diese Annahmen können sich auf unterschiedliche Inhalte beziehen. Anhaltspunkte dafür, welche Merkmale Lehrkräfte für Beurteilungen von Schülerinnen und Schülern als bedeutsam erachten, liefern zum Beispiel Interview- und MouseLab-Studien, die Übergangsempfehlungen untersuchten. Neben schulischen Leistungen der Heranwachsenden sprechen Lehrkräfte diesen Studien zufolge der Motivation und dem Arbeitsverhalten (z. B. Ausdauer/ Fleiß, Interesse/ Motivation) die höchste Relevanz zu. Zusätzlich griffen die Lehrkräfte auf Informationen zur elterlichen Unterstützung zurück. Weiteren Merkmalen wie beispielsweise dem Sozialverhalten der Kinder oder dem Beruf der Eltern wurde eine vergleichsweise geringere Bedeutung zugesprochen (Nölle et al. 2009; Böhmer et al. 2015). Dass die Motivation und das Arbeitsverhalten von Lernenden tatsächlich prädiktiv für Schulleistungen und weiteres Lernen sind (vgl. z. B. Spinath 2011), spricht ebenfalls für die Annahme, dass Lehrkräfte motivationalen Merkmalen bei der Prognose von Leistungsentwicklungen hohe Relevanz beimessen und sie in ihre Leistungserwartungen einbeziehen. Auch Studienergebnisse zu Zusammenhängen von Leistungserwartungen und Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens legen nahe, dass dieser

Einbezug tatsächlich stattfindet: Ein Kind, dessen Motivation und Arbeitsverhalten von einer Lehrkraft als vergleichsweise günstig eingeschätzt wird, scheint bei gleichen Leistungen auch als begabter und leistungsstärker wahrgenommen zu werden (Schrader und Helmke 1990; Anders et al. 2010), bessere Noten zu erhalten (Trautwein und Baeriswyl 2007; Anders et al. 2010; Kuhl und Hannover 2012) und eher für den Übergang auf ein Gymnasium empfohlen zu werden (Anders et al. 2010).

Legt man auch für die Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens von Schülerinnen und Schülern das Kontinuum-Modell der Personenwahrnehmung zugrunde, existieren zwei Wege, über die die Lehrkräfte zu ihren Einschätzungen gelangen können: Zum einen über die Verarbeitung individueller Informationen, die ihnen über die Kinder vorliegen, und zum anderen über die Anwendung stereotyper Annahmen über die Motivation und das Arbeitsverhalten für bestimmte Gruppen. Eine Verarbeitung anhand von Stereotypen ist durchaus wahrscheinlich, da einige Studien zeigen, dass Lehrkräfte die Lernmotivation (Spinath 2005) sowie verwandte emotional-motivationale Merkmale wie z. B. das schulische Interesse (Karing 2009) oder lernbezogene Emotionen (Givvin et al. 2001) nur relativ ungenau einschätzen. Einschätzungen der Motivation und des Arbeitsverhaltens könnten daher ebenfalls anfällig für Einflüsse stereotyper Personenwahrnehmung sein.

Inwieweit Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens systematisch mit Hintergrundmerkmalen der Schülerinnen und Schüler variieren und ob diese möglichen Variationen Unterschiede in den Selbsteinschätzungen der Lernenden widerspiegeln, soll im Folgenden dargestellt werden.

2.3 Unterschiede in Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund und Geschlecht der Schülerinnen und Schüler

Bislang haben nur wenige Studien untersucht, ob Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens von Lernenden systematisch mit Hintergrundmerkmalen wie der ethnischen und sozialen Herkunft oder dem Geschlecht der Schülerinnen und Schüler in Verbindung stehen und die entsprechenden Befunde sind uneinheitlich. Außerdem wurden Selbsteinschätzungen der Schülerinnen und Schüler in den meisten Studien nicht berücksichtigt. Nur drei Studien liefern bislang Anhaltspunkte dafür, inwieweit die Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens den Selbsteinschätzungen der Schülerinnen und Schüler entsprechen oder ob sie eher auf stereotypen Annahmen über diese Merkmale basieren.

Bezogen auf den *ethnischen Hintergrund* Heranwachsender zeigte eine experimentelle Studie aus den USA, dass Grundschullehrkräfte Jungen mit kaukasisch-amerikanisch klingendem Namen als schulisch motivierter beschrieben als Jungen mit afro-amerikanisch klingendem Namen (Anderson-Clark et al. 2008). In einer ebenfalls in den USA durchgeführten Feldstudie schätzten Lehrkräfte Heranwachsende mit und ohne Zuwanderungshintergrund dagegen ähnlich anstrengungsbereit ein, was mit den Selbsteinschätzungen der Schülerinnen und Schüler übereinstimmte (Jussim et al. 1996).

In Deutschland wurden bisher in zwei Studien Lehrkräfteeinschätzungen zu motivationalen Merkmalen und Aspekten des Arbeitsverhaltens für Heranwachsende mit und ohne Zuwanderungshintergrund analysiert. In diesen Studien nahmen Lehrkräfte für Kinder ohne Zuwanderungshintergrund eine etwas höhere Motivation und ein etwas günstigeres Arbeitsverhalten (Anders et al. 2010) sowie eine leicht höhere Schulfreude, Anstrengungsbereitschaft und günstigere schulische Arbeitshaltung an (Schneider 2011) als für Kinder aus zugewanderten Familien, wobei jedoch keine Selbsteinschätzungen der Kinder kontrolliert wurden. Ergebnisse für verschiedene ethnische Gruppen in Deutschland liegen bisher ebenfalls nicht vor.

Auch in Abhängigkeit von der *sozialen Herkunft* der Heranwachsenden identifizierten verschiedene Studien Unterschiede in Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens. So wurden Kinder aus sozial bessergestellten Familien nicht nur in Deutschland (Trautwein und Baeriswyl 2007; Anders et al. 2010) sondern auch in Studien aus den USA (Jussim et al. 1996) tendenziell als motivierter und anstrengungsbereiter eingeschätzt als Kinder aus sozial benachteiligten Familien. Diese von den Lehrkräften angenommenen Gruppenunterschiede fielen dabei stärker aus als die entsprechenden Gruppenunterschiede in den Selbsteinschätzungen der Kinder (Jussim et al. 1996).

Darüber hinaus weisen mehrere Studienergebnisse darauf hin, dass Lehrkräfte die Motivation und das Arbeitsverhalten in Abhängigkeit vom *Geschlecht* der Kinder unterschiedlich einschätzen. So scheinen Lehrkräfte davon auszugehen, dass Mädchen im Vergleich zu Jungen insgesamt motivierter sind (Trautwein und Baeriswyl 2007; Anders et al. 2010; Schneider 2011), sich stärker anstrengen (Jussim und Eccles 1992; Jussim et al. 1996) sowie über eine höhere schulische Selbstregulationskompetenz (Duckworth und Seligman 2006; Kuhl und Hannover 2012) verfügen. Inwieweit diese angenommenen Geschlechtsunterschiede Differenzen in Selbsteinschätzungen der jeweiligen Schülerinnen und Schüler widerspiegeln, bleibt allerdings uneindeutig. Lediglich drei Studien haben zusätzlich zu den Einschätzungen der Lehrkräfte auch Selbsteinschätzungen der Schülerinnen und Schüler berücksichtigt. Diese Studien kamen zu uneinheitlichen Ergebnissen: In der Studie von Duckworth und Seligman (2006) stimmten die von den Lehrkräften angenommenen Geschlechtsunterschiede in der Selbstregulationskompetenz mit den Selbsteinschätzungen der Mädchen und Jungen überein. In den Studien von Jussim und Eccles (1992) und Jussim et al. (1996) wurden hingegen Differenzen zwischen den Einschätzungen der Lehrkräfte und den Angaben der Mädchen und Jungen dazu deutlich, wie sehr sie sich in der Schule anstrengen und wie viel Zeit sie in Hausaufgaben investieren.

In der Zusammenschau deutet die Befundlage darauf hin, dass Lehrkräfte die Motivation und das Arbeitsverhalten von Schülerinnen und Schülern in Abhängigkeit von deren Hintergrundmerkmalen unterschiedlich einschätzen. Da sich diese Unterschiede ersten Vergleichen zufolge nur teilweise in den Selbsteinschätzungen der Kinder widerzuspiegeln scheinen, könnte die Wahrnehmung dieser Unterschiede teilweise auf stereotypen Annahmen der Lehrkräfte beruhen. Aufgrund der unzureichenden und uneinheitlichen Befundlage bedarf diese Hypothese jedoch weiterer empirischer Analysen.

Sollten Lehrkräfte Unterschiede in der Motivation und dem Arbeitsverhalten für verschiedene Gruppen von Schülerinnen und Schülern in Abhängigkeit von deren

ethnischem Hintergrund, sozialem Hintergrund oder Geschlecht annehmen, die sich nicht in deren tatsächlichen Motivlagen und Verhalten widerspiegeln, kann es dadurch zu den beobachteten Verzerrungen in Leistungserwartungen kommen.

2.4 Vermittlung von Verzerrungen in Leistungserwartungen durch die Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens

Die Frage, ob die von Lehrkräften eingeschätzte Motivation und das eingeschätzte Arbeitsverhalten der Schülerinnen und Schüler ethnische, soziale und geschlechtsbezogene Verzerrungen in ihren Leistungserwartungen erklären können, wurde bislang noch nicht empirisch untersucht. Zwei Studien haben allerdings die Mediation des Zusammenhangs zwischen Geschlecht und Noten (als Indikatoren für Leistungseinschätzungen) durch die von der Lehrkraft eingeschätzte Selbstregulationskompetenz der Kinder analysiert und bieten damit erste Hinweise. Die Studien zeigen, dass die von den Lehrkräften als höher eingeschätzte Selbstregulationskompetenz von Mädchen den Zusammenhang zwischen Geschlecht und Noten vermittelte (Duckworth und Seligman 2006; Kuhl und Hannover 2012). In getrennten Analysen für die Deutsch- und Mathematiknoten wurden zudem unterschiedliche Mediationsmuster in den Fächern deutlich (Kuhl und Hannover 2012): Mädchen erhielten im Fach Deutsch auch bei gleicher Leseleistung bessere Noten als Jungen, wobei dieser Zusammenhang größtenteils durch die von den Lehrkräften eingeschätzte Selbstregulationskompetenz vermittelt wurde. Im Gegensatz hierzu waren die Noten der Mädchen und Jungen im Fach Mathematik allgemein und auch unter Kontrolle der Testleistung vergleichbar. Wurde jedoch zusätzlich die Lehrkrachteinschätzung der Selbstregulationskompetenz berücksichtigt, erhielten Jungen im Durchschnitt bessere Mathematiknoten als Mädchen.

Diese Befunde deuten darauf hin, dass Geschlechtsunterschiede in der Benotung durch Lehrkräfte auf deren Annahmen zur Selbstregulationskompetenz, die eine wichtige Voraussetzung für effektives Arbeitsverhalten darstellt, zurückgeführt werden können. Unklar ist allerdings, ob auch ethnische, soziale und geschlechtsbezogene Verzerrungen in Leistungserwartungen über die Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens vermittelt werden.

3 Forschungsfragen und Hypothesen

Der vorliegende Beitrag knüpft an den Analysen von Lorenz et al. (2016) an, in denen anhand von Daten des KuL-Projekts gezeigt werden konnte, dass die Leistungserwartungen von Lehrkräften in Abhängigkeit vom ethnischen Hintergrund, sozialen Hintergrund und Geschlecht der Kinder verzerrt sind. Anhand desselben Datensatzes gehen die im Folgenden beschriebenen Analysen der Frage nach, ob sich diese Verzerrungen der Leistungserwartungen auf Unterschiede in der Lehrkrachteinschätzung der schulischen Motivation und des Arbeitsverhaltens ihrer Schülerinnen und Schüler zurückführen lassen.

Dazu wird zunächst untersucht, ob sich die Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens in Abhängigkeit von den Hintergrundmerkmalen der

Kinder allgemein und nach Kontrolle der Selbsteinschätzungen der Lernenden unterscheiden. Da der theoretische und empirische Forschungsstand zu ethnischen Unterschieden nicht ausreicht, um konkrete Annahmen für einzelne Herkunftsgruppen abzuleiten, wird die entsprechende Hypothese ungerichtet formuliert.

Hypothese 1a Lehrkräfte schätzen die Motivation und das Arbeitsverhalten von Kindern in Abhängigkeit vom ethnischen Hintergrund unterschiedlich ein.

Hypothese 1b Lehrkräfte schätzen die Motivation und das Arbeitsverhalten von Kindern aus sozial bessergestellten Familien positiver ein als von Kindern aus sozial schwächeren Familien.

Hypothese 1c Lehrkräfte schätzen die Motivation und das Arbeitsverhalten von Mädchen positiver ein als von Jungen.

Es wird erwartet, dass die in den Hypothesen 1a–c postulierten Unterschiede auch nach Kontrolle der Selbsteinschätzungen der Motivation und des Arbeitsverhaltens der Schülerinnen und Schüler bestehen bleiben.

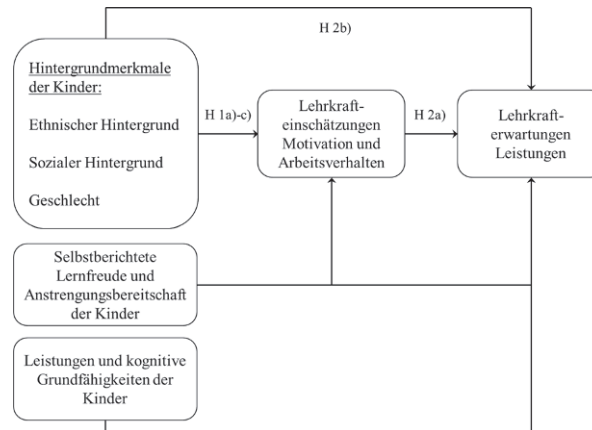
Weiterhin wird der Frage nachgegangen, ob Unterschiede in der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens Verzerrungen in den Leistungserwartungen erklären.

Hypothese 2a Zwischen Leistungserwartungen von Lehrkräften und ihrer Einschätzung der Motivation und des Arbeitsverhaltens der Kinder besteht ein positiver Zusammenhang.

Hypothese 2b Die Zusammenhänge zwischen den Hintergrundmerkmalen der Lernenden und den Leistungserwartungen der Lehrkräfte lassen sich zum Teil auf die Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens zurückführen, d. h. die Koeffizienten des ethnischen Hintergrunds, sozialen Hintergrunds und Geschlechts in Prädiktion der Leistungserwartungen reduzieren sich nach Kontrolle der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens ihrer Schülerinnen und Schüler signifikant.

Eine grafische Darstellung des vollständigen Analysemodells ist in Abb. 1 zu finden.

Abb. 1 Grafische Darstellung des vollständigen Analysemodells und der Hypothesen (H 1a bis H 2b)



4 Methodisches Vorgehen

4.1 Stichprobe

Die vorliegende Untersuchung basiert auf Daten der Studie „Kompetenzerwerb und Lernvoraussetzungen (KuL)“¹, an der im Schuljahr 2013/2014 erste Grundschulklassen aus $N=38$ Schulen aus Nordrhein-Westfalen (NRW) auf freiwilliger Basis teilnahmen. Die Rücklaufquoten der Einwilligungen zur Studienteilnahme lagen auf Schulebene bei 8 %, auf Lehrkräfebene bei 74 % und auf Eltern/ Kind-Ebene bei 66 %. In die folgenden Analysen gehen Daten von insgesamt $N=901$ Kindern aus $N=66$ ersten Klassen mit 69 Klassen- und Fachlehrkräften ein.

Die Lehrkräfte waren im Durchschnitt 41 Jahre und 8 Monate alt ($SD=8,89$) und seit durchschnittlich 12 Jahren ($SD=9,00$) im Lehrberuf tätig. Es handelte sich fast ausschließlich um weibliche Lehrkräfte (94 %) ohne Zuwanderungshintergrund (94 %).

Zum Zeitpunkt der Einschulung waren die Kinder (48 % weiblich) im Mittel 6 Jahre und 5 Monate alt ($SD=0,32$). Laut Elternaussagen zum Geburtsland beider Elternteile sowie der Großeltern lag der Anteil von Kindern mit Zuwanderungshintergrund bei 38 % (mindestens ein Großeltern- oder Elternteil nicht in Deutschland geboren). Kinder der ersten Zuwanderungsgeneration ($N=21$)² wurden ebenso wie Kinder ohne Angaben zum Zuwanderungshintergrund ($N=134$) aus den Analysen ausgeschlossen. Die Angaben zum Herkunftsland der Familien wurden zu insgesamt vier Herkunftsregionen zusammengefasst: „Deutschland“ (61,6 %), „Türkei“ (11,3 %), „ehem. SU und Osteuropa“ (10,9 %) und „andere Länder“ (16,2 %), wobei

¹ Das interdisziplinäre Forschungsprojekt wurde unter der Leitung von Prof. Dr. Cornelia Kristen (Otto-Friedrich-Universität Bamberg), Prof. Dr. Irena Kogan (Universität Mannheim) und Prof. Dr. Petra Stanat (Humboldt-Universität zu Berlin) durchgeführt.

² Die selbst zugewanderten Erstklässlerinnen und Erstklässler leben erst relativ kurz in Deutschland. Es ist anzunehmen, dass dies Urteile von Lehrkräften in besonderer Weise beeinflusst. Separate Analysen wären wünschenswert, waren durch die geringe Fallzahl jedoch nicht möglich.

die Region „chem. SU und Osteuropa“³ alle Staaten der ehemaligen Sowjetunion sowie Polen, Tschechien, Bulgarien, Rumänien, Ungarn, Slowakei und Slowenien umfasst.

Als Indikator des sozioökonomischen Status (SES) dient der *Highest International Socio-Economic Index of Occupational Status* (HISEI; Ganzeboom 2010). Die Familien wiesen durchschnittlich einen HISEI von 53,25 ($SD=19,48$) auf. Verglichen mit dem durchschnittlichen HISEI von 48,10 ($SD=14,80$) für Viertklässlerinnen und Viertklässler in NRW, der im IQB-Ländervergleich 2011 ermittelt wurde (Richter et al. 2012), ist die Stichprobe leicht positiv selektiert.

4.2 Datenerhebung und Messinstrumente

In der dritten bis zwölften Unterrichtswoche⁴ nach Schuljahresanfang gaben die Lehrkräfte ihre Einschätzungen der Leistungen, der Motivation und des Arbeitsverhaltens der teilnehmenden Kinder ihrer Klasse in einem Fragebogen an. Im selben Zeitraum wurden die Kompetenzen der Schülerinnen und Schüler in zwei Einzelsitzungen je Kind (jeweils ca. 30 min) von geschulten Testleiterinnen und Testleitern erfasst.

4.2.1 Leistungserwartungen und Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens

Die Lehrkräfte schätzten für jedes Kind die voraussichtliche Leistungsentwicklung im Verlauf des kommenden Schuljahres sowie die zukünftigen Leistungen am Ende des Schuljahres im sprachlichen und mathematischen Bereich auf einer fünfstufigen Antwortskala im Vergleich zu den anderen Kindern der Klasse ein (z. B. „Wo wird das Kind mit seinen Leistungen am Ende des kommenden Schuljahres im Vergleich zu den anderen Kindern aus dieser Klasse voraussichtlich stehen? a) in Mathematik, b) in Deutsch“).⁵ Die Erwartungen für den mathematischen Bereich wurden mit zwei Items ($\alpha=0,94$), die Erwartungen für den sprachlichen Bereich mit drei Items ($\alpha=0,94$) erfasst. Die Items wurden teils aus dem Projekt BiKS-3-10 (Artelt et al. 2013) adaptiert und teils in der KuL-Studie selbst entwickelt.

Zusätzlich wurden die Lehrkräfte gebeten, die Motivation und das Arbeitsverhalten der Kinder auf einer fünfstufigen Skala (1 = *trifft überhaupt nicht zu* bis 5 = *trifft voll und ganz zu*) einzuschätzen. Auch diese Items wurden teils aus dem Projekt BiKS-3-10 (Artelt et al. 2013) adaptiert und teils in der KuL-Studie selbst entwickelt. Von den insgesamt acht Items ($\alpha=0,96$) beziehen sich jeweils drei auf den Deutsch- und Mathematikunterricht (Deutsch: $\alpha=0,92$, Mathematik: $\alpha=0,92$; z. B. „Das Kind hat große Freude am Lernen im Deutsch-/Mathematikunterricht.“) und

³ Zur besseren Lesbarkeit wird nachfolgend für diese Herkunftsregion von „osteuropäischer Hintergrund“ gesprochen. Kinder, deren Familien aus einem nicht-europäischen Land der ehemaligen Sowjetunion zugewandert sind, sind damit gleichermaßen gemeint.

⁴ Der Zeitpunkt, zu dem die Lehrkräfte den Fragebogen zurückschickten, hing nicht mit der Genauigkeit der leistungsbezogenen Erwartungen der Lehrkräfte zusammen (Lorenz et al. 2016).

⁵ Für den genauen Wortlaut aller Items und der Antwortskalen siehe Abb. A1 im Online Anhang.

zwei allgemein auf das Lernen in der Schule (z. B. „Das Kind ist sehr fleißig in der Schule.“). Die Items decken verschiedene motivationale Aspekte und Aspekte des Arbeitsverhaltens ab (vgl. Abb. A1 im Online Anhang). Zur Erhöhung der Vergleichbarkeit zwischen den Lehrkräfteeinschätzungen und den Selbsteinschätzungen der Kinder, die sich fast ausschließlich allgemein auf das Lernen in der Schule beziehen und nicht nach Fächern differenzieren (vgl. Abschn. 4.2.3), wurden die Items in allen dargestellten Analysen als Gesamtmittelwert berücksichtigt. Um die Robustheit der Ergebnisse abzusichern, wurden alle Analysen jedoch auch mit nach Fächern getrennten Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens durchgeführt. Die Ergebnisse fallen sehr ähnlich aus und werden daher nicht dargestellt. Auf bedeutsam abweichende Ergebnisse wird jeweils hingewiesen.

4.2.2 Sprachliche, mathematische und allgemeine kognitive Schülerfähigkeiten

Die mathematischen und sprachlichen (Vorläufer-)Kompetenzen der Kinder wurden mit dem computergestützten Instrument *Fähigkeitsindikatoren Primarschule* (FIPS; Bäuerlein et al. 2012) erhoben. Die Skala *Lautbewusstheit* ($\alpha = 0,82$) erfasst Fähigkeiten der phonologischen Bewusstheit. Sie setzt sich aus 26 Aufgaben zusammen, bei denen die Kinder Silben und Fantasiewörter nachsprechen, Reime erkennen und Wörter in ihre einzelnen Laute zerlegen sollen. Die Skala *Lesen* ($\alpha = 0,96$) umfasst 88 Aufgaben, die Vorstellungen des Kindes über das Lesen sowie erste Leseaufgaben beinhalten. Der Skala *Mathematik* ($\alpha = 0,92$) sind 53 Aufgaben zugeordnet. Nach leichteren Additions- und Subtraktionsaufgaben anhand von Bildern und dem Vorlesen verschiedener Zahlen, erhöht sich der Abstraktionsgrad über Rechenaufgaben anhand von Punkten bis hin zu formalen Rechenaufgaben. Die im Schwierigkeitsgrad aufsteigenden Aufgaben wurden den Kindern in allen Kompetenzbereichen solange dargeboten, bis sie eine bestimmte Anzahl an falschen Antworten gegeben haben. Je nach Kompetenzstand erfasst das Instrument damit bei einigen Kindern vor allem Vorläuferkompetenzen, bei anderen Kindern, die in der Entwicklung bereits fortgeschritten sind, auch sprachliche und mathematische Kompetenzen.

Zusätzlich wurden kognitive Grundfähigkeiten der Kinder erfasst: Zum einen mit dem *Matrizentest* des *Grundintelligenztests Skala 1* (CFT1; Weiß und Osterland 1997; $\alpha = 0,78$) als Maß des schlussfolgernden Denkens und zum anderen mit der Unterskala *Kurzzeitgedächtnis* des FIPS (Bäuerlein et al. 2012; $\alpha = 0,76$) als Maß des räumlich-visuellen Arbeitsgedächtnisses.

4.2.3 Schülerelbsteinschätzung der Motivation und des Arbeitsverhaltens

In der Mitte des Schuljahres wurden die Kinder zu ihrer Motivation und zu ihrem Arbeitsverhalten befragt. Dazu wurden zwei Unterskalen des FEES1-2 (Rauer und Schuck 2004) in adaptierter Version eingesetzt. Die Kinder machten auf einer dreistufigen Skala (0 = *trifft nicht zu* bis 2 = *trifft zu*) Angaben zu ihrer Lernfreude ($\alpha = 0,78$, 13 Items; z. B. „Ich lerne gern in der Schule.“) und Anstrengungsbereitschaft ($\alpha = 0,70$, 13 Items; z. B. „Ich versuche, auch ganz schwierige Aufgaben zu lösen.“). Die Items beziehen sich größtenteils allgemein auf das Lernen in der Schule.

4.3 Analysestrategie

Zunächst wurde mit Regressionsanalysen untersucht, ob sich die Lehrkrachteinschätzungen der Motivation und des Arbeitsverhaltens in Abhängigkeit von den Hintergrundmerkmalen der Kinder (ethnischer Hintergrund, sozialer Hintergrund, Geschlecht) allgemein unterscheiden (Hypothesen 1a–c). Um spezifische Effekte identifizieren zu können, erfolgte die Aufnahme aller drei Merkmale in das Modell simultan (Tab. 3, Modell 1). Um zu überprüfen, inwieweit die gefundenen Unterschiede mit den tatsächlichen Motivlagen und Arbeitshaltungen der Kinder übereinstimmen oder darüber hinausgehen, wurden die Selbsteinschätzungen der Lernfreude und Anstrengungsbereitschaft in einem nächsten Schritt hinzugenommen (Tab. 3, Modell 2).

Mit Pfadmodellen wurde anschließend analysiert, inwieweit sich Verzerrungen in Leistungserwartungen auf Unterschiede in der Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens der Kinder zurückführen lassen (Hypothesen 2a–b). Als Verzerrungen werden hierbei ethnische, soziale und geschlechtsbezogene Unterschiede in den Leistungserwartungen verstanden, die nach Kontrolle fachspezifischer Leistungen der Kinder, ihrer kognitiven Grundfähigkeiten und ihrer selbsteingeschätzten Lernfreude und Anstrengungsbereitschaft bestehen bleiben (Tab. 4 und 5, Modell 0). Um den Anteil zu bestimmen, mit dem sich Verzerrungen in Leistungserwartungen auf die Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens zurückführen lassen, wurden die Regressionskoeffizienten der Hintergrundmerkmale vor und nach zusätzlicher Kontrolle der Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens miteinander verglichen und indirekte Pfade spezifiziert (Tab. 4 und 5, Modell 1).

Die Modellschätzungen erfolgten mit der Software Mplus Version 7.3 (Muthén und Muthén 1998–2012). Die Schätzung der Standardfehler wurde für die Mehrebenenstruktur der Daten korrigiert und die akzeptierte α -Fehlerwahrscheinlichkeit auf 5 % festgelegt. Alle metrischen Prädiktorvariablen gingen z-standardisiert in die Analysen ein. Die Regressionskoeffizienten wurden an der Varianz der abhängigen Variable standardisiert (y-Standardisierung). Zum Umgang mit fehlenden Werten wurde das in Mplus implementierte Full-Information-Maximum-Likelihood-Verfahren (FIML) unter Einbezug vielfältiger Hilfsvariablen angewendet, u. a. zusätzliche Lehrkrachteinschätzungen, Leistungsmaße sowie Merkmale des familiären Lernumfelds. Der Anteil fehlender Werte in den Analysevariablen lag bei maximal 6,9 % (vgl. Tab. A1 im Online Anhang).

5 Ergebnisse

5.1 Deskriptive Ergebnisse

Mittelwerte und Standardabweichungen der Leistungserwartungen und der Lehrkrachteinschätzung der Motivation und des Arbeitsverhaltens sowohl für die Analysetichprobe insgesamt als auch für die verschiedenen Schülergruppen sind in Tab. 1 zu finden. Tab. 2 enthält Angaben zu den Leistungen und zur selbsteingeschätzten

Lernfreude und Anstrengungsbereitschaft der Lernenden. Die zentralen Ergebnisse sollen im Folgenden kurz beschrieben werden.

Die Mittelwerte der Leistungserwartungen lagen sowohl für den sprachlichen Bereich ($M=3,39$; $SD=0,94$) als auch für den mathematischen Bereich ($M=3,43$; $SD=0,85$) insgesamt knapp oberhalb des theoretischen Mittelpunkts der Skala. Die Lehrkräfte tendierten also insgesamt zu leicht positiven Erwartungshaltungen. Sie gingen demnach davon aus, dass die an der Studie teilnehmenden Kinder ihrer Klasse durchschnittlich eine etwas bessere Leistungsentwicklung zeigen werden als die Kinder der Klasse insgesamt. Bei der Einschätzung der Motivation und des Arbeitsverhaltens war die Tendenz zu Urteilen oberhalb des theoretischen Mittelpunkts der Skala noch etwas stärker ausgeprägt ($M=3,71$; $SD=0,85$). Sowohl die Leistungserwartungen als auch die Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens unterschieden sich in Abhängigkeit von den Hintergrundmerkmalen der Lernenden (vgl. Tab. 1). So fielen beispielsweise die Leistungserwartungen für türkischstämmige Kinder sowohl im sprachlichen als auch im mathematischen Bereich geringer aus als für Kinder ohne Zuwanderungshintergrund. Ihre Motivation und ihr Arbeitsverhalten beurteilten die Lehrkräfte aber genauso hoch wie von Kindern ohne Zuwanderungshintergrund. Außerdem schätzten die Lehrkräfte alle berücksichtigten Merkmale der Lernenden umso günstiger ein, je höher der soziale Status der Familien der Kinder war. Ebenfalls sprachen sie Mädchen eine bessere Leistungsentwicklung im sprachlichen Bereich, eine höhere Motivation und ein günstigeres Arbeitsverhalten zu als Jungen, wohingegen sie für Jungen einen stärkeren Leistungszuwachs in Mathematik erwarteten als für Mädchen.

Mit Blick auf die Testleistungen der Schülerinnen und Schüler (vgl. Tab. 2) kann festgehalten werden, dass die Unterschiede in den Leistungserwartungen zum Teil objektiv beobachtbare Leistungsunterschiede zwischen den Gruppen widerspiegeln. Teilweise ließen sich jedoch keine entsprechenden Unterschiede in den Testleistungen der Kinder identifizieren. Die Selbsteinschätzungen der Motivation und des Arbeitsverhaltens der Lernenden unterschieden sich vergleichsweise wenig zwischen den Gruppen. Lediglich Kinder aus sozial bessergestellten Familien gaben tendenziell eine höhere Anstrengungsbereitschaft an als Kinder aus sozial schwächeren Familien. Außerdem fiel die selbsteingeschätzte Lernfreude der Mädchen etwas höher aus als die der Jungen.

Um zu prüfen, ob sich die Befunde zur Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens domänenspezifisch unterscheiden, wurden die fächerbezogenen Items jeweils auch getrennt analysiert (vgl. Tab. 1). Dabei zeigte sich, dass die positivere Einschätzung der Motivation und des Arbeitsverhaltens von Mädchen bezogen auf den Deutschunterricht (Mädchen: $M=3,87$, $SD=0,87$; Jungen: $M=3,53$, $SD=0,84$, $t(871)=-5,82$, $p<0,05$) stärker ausfiel als bezogen auf das Fach Mathematik (Mädchen: $M=3,79$, $SD=0,89$; Jungen: $M=3,61$, $SD=0,88$, $t(838)=-2,97$, $p<0,05$). Bezüglich des ethnischen und sozialen Hintergrunds der Kinder waren keine Unterschiede zwischen den Domänen erkennbar. Eine Übersicht aller Interkorrelationen der Analyse- und Kontrollvariablen ist in Tab. A1 im Online Anhang zu finden.

Tab. 1 Mittelwerte und Standardabweichungen der Leistungserwartungen und der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens für verschiedene Schülergruppen

	Leistungserwartungen		Lehrkräfteeinschätzung Motivation und Arbeitsverhalten		
	Sprache	Mathe	Gesamt	Deutsch	Mathe
<i>Gesamt</i>	3,39 (0,94)	3,43 (0,85)	3,71 (0,85)	3,69 (0,87)	3,70 (0,89)
<i>Ethnischer Hintergrund</i>					
Deutschland	3,49 ^{c,d} (0,92)	3,46 ^{g,h} (0,85)	3,73 (0,85)	3,73 (0,86)	3,72 (0,89)
Türkei	2,98 ^{c,e} (0,96)	3,16 ^{g,i} (0,89)	3,55 ^k (0,87)	3,52 ^m (0,93)	3,50 ^o (0,92)
SU/Osteuropa	3,61 ^{e,f} (0,86)	3,76 ^{h,i,j} (0,79)	3,91 ^{k,l} (0,79)	3,88 ^{m,n} (0,81)	3,89 ^o (0,86)
Andere Länder	3,15 ^{d,f} (0,91)	3,29 ^j (0,82)	3,61 ^l (0,84)	3,56 ⁿ (0,87)	3,62 (0,88)
<i>Sozialer Hintergrund^a</i>					
Niedrig	3,00 (0,95)	3,11 (0,85)	3,45 (0,79)	3,42 (0,82)	3,40 (0,82)
Mittel	3,42 (0,91)	3,44 (0,83)	3,73 (0,87)	3,71 (0,89)	3,73 (0,91)
Hoch	3,75 (0,82)	3,75 (0,78)	3,95 (0,78)	3,97 (0,80)	3,96 (0,83)
<i>Geschlecht^b</i>					
Mädchen	3,48 (0,96)	3,34 (0,86)	3,88 (0,84)	3,87 (0,87)	3,79 (0,89)
Junge	3,31 (0,91)	3,51 (0,85)	3,55 (0,82)	3,53 (0,84)	3,61 (0,88)

Games-Howell-Test für ungleiche Varianzen (ethnischer und sozialer Hintergrund) bzw. t-Test (Geschlecht)

^aAlle Mittelwertsunterschiede zwischen den sozialen Gruppen auf dem 5 %-Niveau signifikant

^bAlle Mittelwertsunterschiede zwischen den Geschlechtern auf dem 5 %-Niveau signifikant

^{c-o}Mittelwertsunterschied zwischen den jeweiligen Gruppen auf dem 5 %-Niveau signifikant

5.2 Unterschiede in der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens und Zusammenhänge mit Selbsteinschätzungen der Kinder

Zunächst wurde die Annahme geprüft, dass Lehrkräfte die Motivation und das Arbeitsverhalten von Schülerinnen und Schülern in Abhängigkeit von deren Hintergrundmerkmalen unterschiedlich einschätzen (Hypothesen 1a–c; vgl. Tab. 3).

Unter Kontrolle des SES und Geschlechts schätzten Lehrkräfte Kinder mit osteuropäischem Hintergrund als motivierter ein und sprachen ihnen ein günstigeres Arbeitsverhalten zu als Kindern ohne Zuwanderungshintergrund ($\beta_y = 0,26$, $p < 0,05$). Unterschiede für türkischstämmige Kinder im Vergleich zu Kindern ohne Zuwanderungshintergrund zeigten sich hingegen nicht ($\beta_y = -0,06$, $p = 0,51$). Lehrkräfte schrieben jedoch Kindern eine umso höhere Motivation und ein umso günstigeres

Tab. 2 Mittelwerte und Standardabweichungen der Leistungen sowie der Selbsteinschätzungen der Motivation und des Arbeitsverhaltens verschiedener Schülergruppen

	Leistungen			Selbsteinschätzung Motivation und Arbeitsverhalten	
	Lautbew.	Lesen	Mathe	Lernfreude	Anstreng.
<i>Gesamt</i>	13,54 (4,57)	23,99 (26,99)	27,58 (7,73)	1,54 (0,38)	1,70 (0,28)
<i>Ethnischer Hintergrund</i>					
Deutschland	13,77 ^a (4,56)	23,45 (25,97)	28,35 ^{d,e} (7,52)	1,52 (0,39)	1,70 (0,28)
Türkei	11,83 ^{a,b,c} (4,30)	19,53 (22,28)	24,70 ^{d,f} (7,74)	1,56 (0,35)	1,65 (0,29)
SU/Osteuropa	14,36 ^b (4,45)	29,98 (32,51)	29,07 ^{f,g} (7,54)	1,54 (0,39)	1,73 (0,26)
Andere Länder	13,35 ^c (4,65)	25,47 (29,49)	25,78 ^{e,g} (7,91)	1,58 (0,32)	1,72 (0,25)
<i>Sozialer Hintergrund</i>					
Niedrig	11,70 ^h (4,32)	17,82 ^{i,j} (19,71)	24,78 ^k (8,60)	1,51 (0,37)	1,64 ^{l,m} (0,31)
Mittel	13,67 ^h (4,37)	24,20 ⁱ (27,28)	27,90 ^k (7,10)	1,54 (0,38)	1,71 ^l (0,27)
Hoch	15,36 ^h (4,55)	30,19 ^j (30,78)	30,06 ^k (7,27)	1,57 (0,37)	1,76 ^m (0,23)
<i>Geschlecht</i>					
Mädchen	13,60 (4,51)	23,89 (24,40)	25,61 ⁿ (6,87)	1,59 ^o (0,34)	1,71 (0,26)
Junge	13,48 (4,64)	24,08 (29,13)	29,35 ⁿ (8,04)	1,50 ^o (0,40)	1,70 (0,29)

Games-Howell-Test für ungleiche Varianzen (ethnischer und sozialer Hintergrund) bzw. t-Test (Geschlecht)

^{a-o}Mittelwertsunterschied zwischen den jeweiligen Gruppen auf dem 5%-Niveau signifikant

Arbeitsverhalten zu, je höher der SES ihrer Familien war ($\beta_y = 0,22$, $p < 0,05$). Zudem wurden Mädchen im Vergleich zu Jungen im Durchschnitt als motivierter und anstrengungsbereiter eingeschätzt ($\beta_y = 0,42$, $p < 0,05$). Diese Unterschiede spiegelten sich nur zu einem sehr geringen Teil in der selbsteingeschätzten Lernfreude und Anstrengungsbereitschaft der Kinder wider und blieben auch nach Kontrolle dieser Merkmale mit geringfügig reduzierter Effektstärke bestehen (vgl. Tab. 3, Modell 2).

5.3 Vermittlung der Verzerrungen in Leistungserwartungen durch Lehrkräfteinschätzungen der Motivation und des Arbeitsverhaltens

Als Basis für die Mediationsanalysen wurden zunächst die Verzerrungen in den Leistungserwartungen der Lehrkräfte bestimmt. Neben den Hintergrundmerkmalen der Kinder wurden hierzu die bereichsspezifischen Ausgangsleistungen der Lernenden, ihre kognitiven Grundfähigkeiten sowie ihre selbsteingeschätzte Lernfreude und An-

Tab. 3 Prädiktion der Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens

	Modell 1			Modell 2		
	<i>b</i>	<i>SE b</i>	β_y	<i>b</i>	<i>SE b</i>	β_y
<i>Ethnischer Hintergrund (Referenz = Deutschland)</i>						
Türkei	−0,05	0,07	−0,06	−0,05	0,07	−0,06
SU/Osteuropa	0,22	0,08	0,26*	0,19	0,08	0,23*
Andere Länder	−0,05	0,06	−0,06	−0,09	0,07	−0,10
<i>Sozialer Hintergrund</i>						
Sozioökonomischer Status	0,19	0,03	0,22*	0,16	0,03	0,19*
<i>Geschlecht</i>						
Weiblich	0,35	0,06	0,42*	0,32	0,05	0,37*
<i>Selbsteinschätzung</i>						
Lernfreude	–	–	–	0,12	0,05	0,14*
Anstrengungsbereitschaft	–	–	–	0,14	0,04	0,16*
Intercept	3,53*	0,05	–	3,55*	0,05	–
R^2	10,2 %*			17,2 %*		

* $p < 0,05$

strebungsbereitschaft als Prädiktoren der Leistungserwartungen einbezogen. Trotz der leicht modifizierten Operationalisierung der Leistungserwartungen sind die gefundenen Verzerrungen mit den von Lorenz et al. (2016) berichteten Ergebnissen konsistent (vgl. Tab. 4 und 5, Modell 0). So erwarteten Lehrkräfte unter Kontrolle der Leistungen und selbsteingeschätzten Lernfreude und Anstrengungsbereitschaft der Kinder einen geringeren Leistungszuwachs im sprachlichen Bereich für türkischstämmige Kinder im Vergleich zu Kindern ohne Zuwanderungshintergrund ($\beta_y = -0,32$, $p < 0,05$), für Kinder aus sozial schwächeren Familien im Vergleich zu Kindern aus sozial bessergestellten Familien ($\beta_y = 0,15$, $p < 0,05$) und für Jungen im Vergleich zu Mädchen ($\beta_y = 0,17$, $p < 0,05$). Im mathematischen Bereich fielen die Leistungserwartungen höher aus für Kinder mit osteuropäischem Hintergrund als für Kinder ohne Zuwanderungshintergrund ($\beta_y = 0,34$, $p < 0,05$) und höher für Kinder aus sozial bessergestellten Familien als für Kinder aus sozial benachteiligten Familien ($\beta_y = 0,13$, $p < 0,05$). Geschlechtsbezogene Verzerrungen waren im mathematischen Bereich nicht zu verzeichnen ($\beta_y = -0,07$, $p = 0,34$).

Um die Hypothesen 2a und 2b zu prüfen, wurden die Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens als Mediator zwischen den Hintergrundmerkmalen der Kinder und den Leistungserwartungen hinzugenommen und indirekte Pfade spezifiziert (vgl. Tab. 4 und 5, Modell 1; siehe auch Abb. 1). Die Ergebnisse zeigen, dass je positiver die Lehrkräfte die Motivation und das Arbeitsverhalten der Kinder einschätzten, desto höher fielen insgesamt auch die Leistungserwartungen im sprachlichen und mathematischen Bereich aus (Sprache: $\beta_y = 0,52$, $p < 0,05$; Mathematik: $\beta_y = 0,54$, $p < 0,05$). Dieser Zusammenhang ergab sich bei gleichen Leistungen, bei gleicher selbsteingeschätzter Lernfreude und Anstrengungsbereitschaft sowie bei gleichen Hintergrundmerkmalen der Kinder.

Die Koeffizienten des *ethnischen Hintergrunds* in der Prädiktion der Leistungserwartungen änderten sich nach Kontrolle der Lehrkräfteeinschätzung der Motivation

Tab. 4 Prädiktion der sprachlichen Leistungserwartungen von Lehrkräften

	Modell 0			Modell 1			
	<i>b</i>	<i>SE b</i>	β_y	<i>b</i>	<i>SE b</i>	β_y	$\Delta\beta_y$
<i>Ethnischer Hintergrund (Referenz = Deutschland)</i>							
Türkei	-0,30	0,09	-0,32*	-0,28	0,08	-0,31*	-0,01
SU/Osteuropa	0,07	0,08	0,07	-0,01	0,06	-0,01	0,08
Andere Länder	-0,27	0,09	-0,29*	-0,23	0,08	-0,25*	-0,04
<i>Sozialer Hintergrund</i>							
SES	0,14	0,03	0,15*	0,07	0,03	0,08*	0,07*
<i>Geschlecht</i>							
Weiblich	0,15	0,07	0,17*	-0,02	0,05	-0,02	0,19*
<i>Leistungen</i>							
Lautbewusstheit	0,20	0,04	0,22*	0,13	0,03	0,14*	—
Lesen	0,20	0,03	0,22*	0,16	0,03	0,18*	—
<i>Kognitive Grundfähigkeiten</i>							
Kurzzeitgedächtnis	0,09	0,03	0,10*	0,03	0,02	0,03	—
Matrizentest	0,13	0,03	0,14*	0,07	0,03	0,08*	—
<i>Selbsteinschätzung</i>							
Lernfreude	0,07	0,04	0,08*	0,01	0,03	0,02	—
Anstrengungsbereitschaft	0,07	0,03	0,07*	0,02	0,03	0,02	—
<i>Lehrkräfteeinschätzung</i>							
Motivation/Arbeitsverhalten	—	—	—	0,49	0,04	0,52*	—
<i>Indirekte Effekte über Lehrkräfteeinschätzung Motivation/Arbeitsverhalten</i>							
Türkei	—	—	—	0,00	0,04	0,00	—
SU/Osteuropa	—	—	—	0,09	0,04	0,10*	—
Andere Länder	—	—	—	-0,03	0,04	-0,03	—
SES	—	—	—	0,10	0,02	0,11*	—
Weiblich	—	—	—	0,19	0,03	0,20*	—
Intercept	3,38*	—	—	3,47*	—	—	—
R^2	36,2%*			56,4%*			

$\Delta\beta_y$ = Differenz von β_y zwischen Modell 0 und Modell 1

SES Sozioökonomischer Status

* $p < 0,05$

und des Arbeitsverhaltens weder im sprachlichen noch im mathematischen Bereich (alle $\Delta\beta_y$ nicht signifikant). Demnach scheinen die mit dem ethnischen Hintergrund verbundenen Verzerrungen der Leistungserwartungen unabhängig von der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens zu sein.

Anders stellte sich das Bild für Verzerrungen der Leistungserwartungen in Abhängigkeit vom *sozialen Hintergrund* der Kinder dar. Die Koeffizienten des SES reduzierten sich nach Kontrolle der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens signifikant (Sprache: $\Delta\beta_y = 0,07$, $p < 0,05$; Mathematik: $\Delta\beta_y = 0,08$, $p < 0,05$). Etwa die Hälfte der mit dem SES assoziierten Verzerrungen ließ sich auf Unterschiede in der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens der Kinder zurückführen. Die über diese Lehrkräfteeinschätzung vermittelten

Tab. 5 Prädiktion der mathematischen Leistungserwartungen von Lehrkräften

	Modell 0			Modell 1			
	<i>b</i>	<i>SE b</i>	β_y	<i>b</i>	<i>SE b</i>	β_y	$\Delta\beta_y$
<i>Ethnischer Hintergrund (Referenz = Deutschland)</i>							
Türkei	-0,06	0,06	-0,07	-0,07	0,05	-0,09	0,02
SU/Osteuropa	0,29	0,09	0,34*	0,21	0,06	0,25*	0,09
Andere Länder	0,00	0,08	0,00	0,01	0,06	0,01	-0,01
<i>Sozialer Hintergrund</i>							
SES	0,11	0,03	0,13*	0,05	0,02	0,05*	0,08*
<i>Geschlecht</i>							
Weiblich	-0,06	0,06	-0,07	-0,28	0,05	-0,33*	0,26*
<i>Leistungen</i>							
Mathematik	0,33	0,03	0,39*	0,21	0,03	0,25*	–
<i>Kognitive Grundfähigkeiten</i>							
Kurzzeitgedächtnis	0,04	0,02	0,05*	0,00	0,02	0,00	–
Matrizentest	0,13	0,03	0,15*	0,09	0,03	0,11*	–
<i>Selbsteinschätzung</i>							
Lernfreude	0,08	0,04	0,09*	0,02	0,03	0,03	–
Anstrengungsbereitschaft	0,06	0,04	0,07	0,02	0,03	0,02	–
<i>Lehrkräfteeinschätzung</i>							
Motivation/Arbeitsverhalten	–	–	–	0,45	0,03	0,54*	–
<i>Indirekte Effekte über Lehrkräfteeinschätzung Motivation/Arbeitsverhalten</i>							
Türkei	–	–	–	0,02	0,04	0,03	–
SU/Osteuropa	–	–	–	0,08	0,04	0,10	–
Andere Länder	–	–	–	-0,00	0,04	-0,00	–
SES	–	–	–	0,09	0,02	0,10*	–
Weiblich	–	–	–	0,18	0,03	0,21*	–
Intercept	3,43*	0,04	–	3,54*	0,04	–	–
R^2	36,7 %*			57,7 %*			

 $\Delta\beta_y$ = Differenz von β_y zwischen Modell 0 und Modell 1

SES Sozioökonomischer Status

* $p < 0,05$

indirekten Pfade des SES waren statistisch signifikant (Sprache: $\beta_y = 0,11$, $p < 0,05$; Mathematik: $\beta_y = 0,10$, $p < 0,05$).

Für die *geschlechtsbezogenen* Verzerrungen in den Leistungserwartungen kehrten sich die Effekte um: Ohne Berücksichtigung der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens zeigte sich ein Vorteil zugunsten der Mädchen im sprachlichen Bereich und kein Geschlechtseffekt im mathematischen Bereich (vgl. Tab. 4 und 5, Modell 0). Wurde die Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens zusätzlich kontrolliert, verschwand der Geschlechtseffekt im sprachlichen Bereich ($\beta_y = -0,02$, $p = 0,67$) und im mathematischen Bereich trat ein Vorteil zugunsten der Jungen auf ($\beta_y = -0,33$, $p < 0,05$). Bei gleichen Testleistungen, gleicher selbsteingeschätzter Lernfreude und Anstrengungsbereitschaft der Kinder sowie bei gleicher Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens erwarteten die Lehrkräfte also für Jungen bessere Mathematikleistungen als

für Mädchen. Im sprachlichen Bereich erwarteten Lehrkräfte hingegen bei gleicher Leistung und bei gleicher Selbst- und Fremdeinschätzung der Motivation und des Arbeitsverhaltens ähnliche Leistungen für Mädchen und Jungen. Die über die Lehrkräfteinschätzungen der Motivation und des Arbeitsverhaltens vermittelten indirekten Pfade des Geschlechts waren statistisch signifikant (Sprache: $\beta_y = 0,20$, $p < 0,05$; Mathematik: $\beta_y = 0,21$, $p < 0,05$).

Wurden die fächerspezifischen Lehrkräfteinschätzungen der Motivation und des Arbeitsverhaltens in den Modellen berücksichtigt, veränderten sich die Ergebnisse im sprachlichen Bereich nicht bedeutsam. Auch im mathematischen Bereich blieb die Richtung der Befunde identisch. Allerdings fiel der Vorteil der Jungen nach Kontrolle der fächerspezifischen Lehrkräfteinschätzungen der Motivation und des Arbeitsverhaltens etwas geringer aus ($\beta_y = -0,23$, $p < 0,05$), wodurch auch der indirekte Pfad eine etwas geringere Stärke zeigte ($\beta_y = 0,10$, $p < 0,05$).

6 Diskussion

Der vorliegende Beitrag ging zwei Fragestellungen nach: Erstens, ob Lehrkräfteinschätzungen der Motivation und des Arbeitsverhaltens primär die Selbsteinschätzungen von Schülerinnen und Schülern oder eher stereotypenbasierte Annahmen widerspiegeln, und zweitens, ob ethnische, soziale und geschlechtsbezogene Verzerrungen von Leistungserwartungen auf diese Unterschiede zurückgeführt werden können.

Zunächst wurde deutlich, dass Lehrkräfte die Motivation und das Arbeitsverhalten von Schülerinnen und Schülern in Abhängigkeit von deren Hintergrundmerkmalen unterschiedlich einschätzten. Diese Differenzen spiegelten sich dabei nur marginal in den Selbsteinschätzungen der Kinder wider. Es kann also davon ausgegangen werden, dass es sich teilweise um stereotype Annahmen der Lehrkräfte handelt.

Übereinstimmend mit Befunden bisheriger Studien (z. B. Trautwein und Baeriswyl 2007; Anders et al. 2010) schrieben Lehrkräfte Kindern aus sozial bessergestellten Familien im Vergleich zu Kindern aus sozial schwächeren Familien sowie Mädchen im Vergleich zu Jungen eine höhere Motivation und ein günstigeres Arbeitsverhalten zu. Aus der Berücksichtigung der Selbsteinschätzungen der Kinder in der vorliegenden Studie ergibt sich darüber hinaus ein zusätzlicher Erkenntnisgewinn: Die Befunde verdeutlichen, dass die Einschätzungen der Lehrkräfte gegenüber Kindern aus sozial bessergestellten Familien und Mädchen auch dann noch höher ausfielen, wenn die Selbsteinschätzungen der Kinder berücksichtigt wurden. Die Ergebnisse bestätigen damit die Hypothesen 1b und 1c.

Unterschiede in Abhängigkeit vom ethnischen Hintergrund (Hypothese 1a) fanden sich unter Kontrolle des SES und Geschlechts nur für die Gruppe der Kinder mit osteuropäischem Hintergrund: Lehrkräfte schrieben diesen Schülerinnen und Schülern eine höhere Motivation und ein günstigeres Arbeitsverhalten zu als Kindern ohne Zuwanderungshintergrund, auch dann wenn die Selbsteinschätzungen der Kinder kontrolliert wurden. Dieser Befund steht im Gegensatz zu Studien von Anders et al. (2010) und Schneider (2011), die kleine negative Effekte des allgemeinen Zuwanderungshintergrunds auf die Lehrkräfteinschätzung der Motivation und

des Arbeitsverhaltens identifizierten. Allerdings wurden dort verschiedene ethnische Herkunftsgruppen nicht separat betrachtet, sodass sich entgegengesetzte Verzerrungen für verschiedene Zuwanderergruppen möglicherweise überdeckt haben.

Weiterhin zeigte sich, dass Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens mit höheren Leistungserwartungen assoziiert sind, wobei dieser Zusammenhang in Übereinstimmung mit Hypothese 2a auch unter Kontrolle der Testleistungen und der Selbsteinschätzungen der Kinder bestehen blieb. In die Leistungserwartungen von Lehrkräften flossen also nicht nur die tatsächlichen Leistungen der Lernenden sowie deren bei sich selbst wahrgenommene Lernfreude und Anstrengungsbereitschaft ein, sondern zusätzlich auch Lehrkraftannahmen zur Motivation und zum Arbeitsverhalten der Kinder.

Außerdem bestätigte sich die Annahme, dass Unterschiede in der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens in Abhängigkeit vom ethnischen und sozialen Hintergrund sowie vom Geschlecht der Schülerinnen und Schüler zur Erklärung von Verzerrungen in Leistungserwartungen beitragen (Hypothese 2b). Damit unterstützen die Befunde insgesamt die Vermutung, dass Verzerrungen in Leistungserwartungen von Lehrkräften teilweise auf Annahmen über die Motivation und das Arbeitsverhalten verschiedener Gruppen von Schülerinnen und Schülern basieren.

Die mit dem *sozialen Hintergrund* assoziierten Verzerrungen in Leistungserwartungen konnten etwa zur Hälfte auf Unterschiede in der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens zurückgeführt werden. Lehrkräfte scheinen anzunehmen, dass Kinder aus sozial bessergestellten Familien insgesamt motivierter und anstrengungsbereiter sind und dass sich damit einhergehend auch ihre Leistungen günstiger entwickeln. Der verbleibende verzerrte Anteil könnte z. B. auf die Annahme der Lehrkräfte zurückgehen, Kinder aus sozial bessergestellten Familien erhielten in schulischen Belangen mehr Unterstützung von ihren Eltern (vgl. z. B. Nölle et al. 2009). Zukünftige Studien sollten dieser Annahme nachgehen. Zusätzlich zu Lehrkräfteeinschätzungen der elterlichen Unterstützung sollten dabei auch Elternangaben berücksichtigt werden und als Vergleichsmaß für die Lehrkräfteeinschätzungen dienen.

Für die *geschlechtsbezogenen* Verzerrungen in den Leistungserwartungen kehrten sich die Effekte nach Kontrolle der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens um, sodass im mathematischen Bereich ein Vorteil zugunsten der Jungen und im sprachlichen Bereich kein Geschlechtseffekt auftrat. Demnach scheinen Lehrkräfte die Motivation und das Arbeitsverhalten von Jungen insgesamt als ungünstiger anzunehmen als von Mädchen und ihnen daher im sprachlichen Bereich einen Leistungsnachteil zu attestieren, obwohl sich die sprachlichen Leistungen der Mädchen und Jungen in unserer Studie nicht unterschieden. Im mathematischen Bereich scheinen die Lehrkräfte hingegen die höheren Leistungen, die sie den Jungen eigentlich zutrauen, nach unten zu adjustieren, weil sie Jungen als weniger schulisches motiviert einschätzen und ihnen ein ungünstigeres Arbeitsverhalten zusprechen. Zwar berichteten die Jungen auch tatsächlich eine etwas geringere Lernfreude als die Mädchen, diese Unterschiede wurden von den Lehrkräften jedoch überschätzt. Diese Befunde zu Leistungserwartungen und Geschlecht sind weitgehend konsis-

tent mit den Ergebnissen der bereits beschriebenen Studie von Kuhl und Hannover (2012) zur Benotung von Mädchen und Jungen.

Die mit dem *ethnischen Hintergrund* verbundenen Verzerrungen der Leistungserwartungen traten unabhängig von der Lehrkräfteeinschätzung der Motivation und des Arbeitsverhaltens auf. Dies weist darauf hin, dass Leistungserwartungen von Lehrkräften für Kinder aus zugewanderten Familien mit stereotypen Annahmen über andere lernrelevante Merkmale der Kinder oder ihres familiären Umfelds verknüpft sind. Ein möglicher Erklärungsansatz für die höheren Leistungserwartungen an Kinder mit osteuropäischem Hintergrund im mathematischen Bereich könnte sein, dass Lehrkräfte bei den Eltern dieser Kinder eine besondere Wertschätzung der mathematisch-naturwissenschaftlichen Fächer wahrnehmen (z. B. Walter und Taskinen 2008), was sich möglicherweise auch in höheren mathematischen Leistungserwartungen für diese Kinder niederschlagen könnte. Außerdem könnten geringere Erwartungen im sprachlichen Bereich für türkischstämmige Kinder durch die Annahme geringerer Deutschkenntnisse bedingt sein. So thematisierten z. B. in einer Interviewstudie von Kratzmann und Pohlmann-Rother (2012) alle befragten Erzieherinnen fehlende Deutschkenntnisse türkischstämmiger Kinder, die ihrer Einschätzung nach zu Herausforderungen im Kindergartenalltag führen. Allerdings sollte diesbezüglich bedacht werden, dass in unseren Analysen im sprachlichen Bereich die Lautbewusstheit als Leistungsmaß kontrolliert wurde. Da die Testung auf Deutsch erfolgte, bildet dieses Maß zu einem gewissen Anteil auch die sprachlichen Fähigkeiten der Kinder im Deutschen ab. Dennoch wäre es aufschlussreich, in weiterführenden Studien zusätzlich eine Gruppierung anhand des Sprachhintergrunds der Kinder vorzunehmen um eventuelle Unterschiede der Befunde im Vergleich zu den Ergebnissen zu ethnischen Herkunftsgruppen identifizieren zu können. Zukünftige Studien sollten außerdem klären, auf welche Merkmale sich stereotype Annahmen von Lehrkräften für verschiedene ethnische Herkunftsgruppen neben der Motivation und dem Arbeitsverhalten noch beziehen und ob diese die in der vorliegenden Studie nicht erklärten Anteile ethnischer Verzerrungen in Leistungserwartungen erklären können.

Die Ergebnisse des Vergleichs zwischen Selbst- und Lehrkräfteeinschätzungen der Motivation und des Arbeitsverhaltens legen nahe, dass es sich eher um stereotype Annahmen der Lehrkräfte über diese Merkmale der Lernenden handelt. Dennoch ist es auch möglich, dass die Lehrkräfteeinschätzungen tatsächliche Gruppenunterschiede im Lernverhalten der Kinder widerspiegeln, die die Erstklässlerinnen und Erstklässler selbst nicht beurteilen können. Zukünftige Studien sollten daher systematische Beobachtungen des Lernverhaltens der Kinder durchführen und diese als objektiveren Vergleichsmaßstab für die Lehrkräfteeinschätzungen heranziehen.

Bei der Interpretation der Ergebnisse insgesamt ist außerdem zu beachten, dass die Stichprobe der Schülerinnen und Schüler in Bezug auf ihre soziale Zusammensetzung leicht positiv selektiert war. Zudem haben sich möglicherweise besonders engagierte Lehrkräfte freiwillig für die Studienteilnahme entschieden. Geht man davon aus, dass engagiertere Lehrkräfte besonders bemüht sind, ihre Schülerinnen und Schüler präzise einzuschätzen, hätte die höhere Beteiligungsquote dieser Lehrkräfte an der Studie eher eine Unterschätzung der Verzerrungen in den Leistungserwartungen zur Folge. Der ermittelte Zusammenhang zwischen den Leistungserwartungen

und den Lehrkrachteinschätzungen der Motivation und des Arbeitsverhaltens sollte davon jedoch nicht beeinflusst sein.

Werden nicht erklärte Varianzanteile in Leistungserwartungen als Verzerrungen interpretiert, ist es möglich, dass Messungenauigkeiten in den Leistungsmaßen zu einer Unterschätzung des zutreffenden Anteils der Leistungserwartungen und damit zu einer Überschätzung etwaiger Verzerrungen führen. Trotz des Einbezugs vielfältiger Leistungsmaße lässt sich dies in der vorliegenden Studie nicht ausschließen.

Insgesamt weisen die Befunde darauf hin, wie wichtig es ist, Lehrkräfte für den möglichen Einfluss von Stereotypen auf schulische Einschätzungen und Erwartungen zu sensibilisieren. Dies ist vor allem deshalb relevant, weil verzerrte Leistungserwartungen über Prozesse der selbsterfüllenden Prophezeiung die Leistungsentwicklung Heranwachsender beeinflussen (Jussim und Harber 2005) und damit zu einer Verfestigung von Bildungsungleichheiten beitragen können. Dass es möglich ist, den Einfluss von Stereotypen zu reduzieren, zeigt die sozialpsychologische Literatur (vgl. z. B. Blair 2002). Hierzu ist es jedoch erforderlich, die zugrunde liegenden Mechanismen zu kennen und diesen aktiv entgegen zu wirken. Hierzu hat die vorliegende Studie wichtige Erkenntnisse geliefert.

Literatur

- Anders, Y., McElvany, N., & Baumert, J. (2010). Die Einschätzung lernrelevanter Schülermerkmale zum Zeitpunkt des Übergangs von der Grundschule auf die weiterführende Schule: Wie differenziert urteilen Lehrkräfte? In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Hrsg.), *Der Übergang von der Grundschule in die weiterführende Schule: Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (S. 313–330). Bonn: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung.
- Anderson-Clark, T.N., Green, R.J., & Henley, T.B. (2008). The relationship between first names and teacher expectations for achievement motivation. *Journal of Language and Social Psychology*, 27(1), 94–99.
- Artelt, C., Blossfeld, H.-P., Faust, G., Roßbach, H.-G., & Weinert, S. (2013). Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter (BiKS-3-10) (Version: 2. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz). https://doi.org/10.5159/IQB_BIKS_3_10_v2.
- Bäuerlein, K., Beinicke, A., Berger, N., Faust, G., Jost, M., & Schneider, W. (2012). *FIPS Fähigkeitsindikatoren Primarschule: Ein computerbasiertes Diagnoseinstrument zur Erfassung der Lernausgangslage und der Lernentwicklung von Schulanfängern. Manual*. Göttingen: Hogrefe.
- Baumert, J., Watermann, R., & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6(1), 46–71.
- van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, 47(2), 497–527.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6(3), 242–261.
- Böhmer, I., Hörstermann, T., Gräsel, C., Krolak-Schwerdt, S., & Glock, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung: Welcher Urteilsregel folgen Lehrkräfte? *Journal für Bildungsforschung Online*, 7(2), 59–81.
- Diehl, C., & Fick, P. (2016). Ethnische Diskriminierung im deutschen Bildungssystem. In C. Diehl, C. Hunkler, & C. Kristen (Hrsg.), *Ethnische Ungleichheiten im Bildungsverlauf* (S. 243–286). Wiesbaden: Springer VS.
- Duckworth, A.L., & Seligman, M.E.P. (2006). Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98(1), 198–208.

- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Hrsg.), *Advances in experimental social psychology* (Bd. 23, S. 1–74). San Diego: Academic Press.
- Ganter, S. (1997). *Stereotype und Vorurteile: Konzeptualisierung, Operationalisierung und Messung*. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.
- Ganzeboom, H. B. G. (2010). *A new international socio-economic index (ISEI) of occupational status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007*. Annual Conference of International Social Survey Programme, Lisbon.
- Givvin, K. B., Stipek, D. J., Salmon, J. M., & MacGyvers, V. L. (2001). In the eyes of the beholder: students' and teachers' judgments of students' motivation. *Teaching and Teacher Education*, 17(3), 321–331.
- Hachfeld, A., Anders, Y., Schroeder, S., Stanat, P., & Kunter, M. (2010). Does immigration background matter? How teachers' predictions of students' performance relate to student background. *International Journal of Educational Research*, 49(2–3), 78–91.
- Heinze, A., Herwartz-Emden, L., & Reiss, K. (2007). Mathematikkenntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit. *Zeitschrift für Pädagogik*, 53(4), 562–581.
- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101(3), 662–670.
- Hirschauer, M., & Kullmann, H. (2010). Lehrerprofessionalität im Zeichen von Heterogenität – Stereotype bei Lehrkräften als kollegial zu bearbeitende Herausforderung. In J. Hagedorn, V. Schurt, C. Steber & W. Waburg (Hrsg.), *Ethnizität, Geschlecht, Familie und Schule* (S. 351–373). Wiesbaden: Verlag für Sozialwissenschaften.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, 59(3), 297–313.
- Jussim, L., & Eccles, J. (1992). Teacher expectations II: construction and reflection of student achievement. *Journal of Personality and Social Psychology*, 63(6), 947–961.
- Jussim, L., & Harber, K. (2005). Teacher expectations and self-fulfilling prophecies: knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155.
- Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Hrsg.), *Advances in experimental social psychology* (Bd. 28, S. 281–388). San Diego: Academic Press.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbe- reich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 197–209.
- Karing, C., Matthäi, J., & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25(3), 159–172.
- Kratzmann, J., & Pohlmann-Rother, S. (2012). Ethnische Stereotype im Kindergarten? Erzieherinnenhal- tungen gegenüber Zuwanderern aus der Türkei. *Zeitschrift für Pädagogik*, 58(6), 855–876.
- Kuhl, P., & Hannover, B. (2012). Differenzielle Benotungen von Mädchen und Jungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(3), 153–162.
- Lorenz, G. (2017). *Selbsterfüllende Prophezeiungen in der Schule: Leistungserwartungen von Lehrkräften und Kompetenzen von Kindern mit Zuwanderungshintergrund*. Wiesbaden: Springer VS.
- Lorenz, G., Gentrup, S., Kristen, C., Stanat, P., & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(1), 89–111.
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, 46(3), 235–261.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7. Aufl.). Los Angeles, CA: Muthén & Muthén.
- Nölle, I., Hörstermann, T., Krolak-Schwerdt, S., & Gräsel, C. (2009). Relevante diagnostische Informatio- nen bei der Übergangsempfehlung – die Perspektive der Lehrkräfte. *Unterrichtswissenschaft*, 37(4), 294–310.
- Prenzel, M., Sälzer, C., Klieme, E., & Köller, O. (Hrsg.) (2013). *PISA 2012. Fortschritte und Herausfor- derungen in Deutschland*. Münster: Waxmann.
- Rauer, W., & Schuck, K. D. (2004). *FEESSI-2: Fragebogen zur Erfassung emotionaler und sozialer Schul- erfahrungen von Grundschulkindern erster und zweiter Klassen*. Göttingen: Beltz.

- Ready, D.D., & Wright, D.L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: the role of child background and classroom context. *American Educational Research Journal*, 48(2), 335–360.
- Richter, D., Kuhl, P., & Pant, H. A. (2012). Soziale Disparitäten des Kompetenzerwerbs. In P. Stanat, H. A. Pant, K. Böhme, & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 191–207). Münster: Waxmann.
- Rjosk, C., McElvany, N., Anders, Y., & Becker, M. (2011). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung der basalen Lesefähigkeit ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 58(2), 92–105.
- Schneider, T. (2011). Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschulempfehlung. *Zeitschrift für Erziehungswissenschaft*, 14(3), 371–396.
- Schrader, F.-W. (2010). Diagnostische Kompetenz von Eltern und Lehrern. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 102–108). Weinheim: Beltz.
- Schrader, F.-W., & Helmke, A. (1990). Lassen sich Lehrer bei der Leistungsbeurteilung von sachfremden Gesichtspunkten leiten? Eine Untersuchung zu Determinanten diagnostischer Lehrerurteile. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22(4), 312–324.
- Schrader, F.-W., & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Weinheim: Beltz.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1–2), 85–95.
- Spinath, B. (2011). Lernmotivation. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche* (S. 45–55). Wiesbaden: Springer VS.
- Stahl, N. (2007). Schülerwahrnehmung und -beurteilung durch Lehrkräfte. In H. Ditton (Hrsg.), *Kompetenzaufbau und Laufbahnen im Schulsystem* (S. 171–198). Münster: Waxmann.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: a meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762.
- Trautwein, U., & Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind: Referenzgruppeneffekte bei Übertrittsentscheidungen. *Zeitschrift für Pädagogische Psychologie*, 21(2), 119–133.
- Walter, O., & Taskinen, P. (2008). Naturwissenschaftsbezogene Motivationen und Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund in Deutschland. Der Einfluss der Generationen, der Herkunft und des Elternhauses. In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006* (S. 185–203). Wiesbaden: Verlag für Sozialwissenschaften.
- Weiß, R. H., & Osterland, J. (1997). *Grundintelligenztest Skala 1: CFT 1. Handanweisung* (5. Aufl.). Göttingen: Hogrefe.

Teilstudie 3: Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement

Dieser Beitrag ist in der Zeitschrift *Learning and Instruction* erschienen. Die Referenz lautet:

Gentrup, S., Lorenz, G., Kristen, C. & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, 66. Doi: 10.1016/j.learninstruc.2019.101296

The final publication is available at Elsevier via

<http://dx.doi.org/10.1016/j.learninstruc.2019.101296>



Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement

Sarah Gentrup^{a,*}, Georg Lorenz^b, Cornelia Kristen^c, Irena Kogan^d

^a Institute for Educational Science, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Berlin, Germany

^b Institute for Educational Quality Improvement (IQB), Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Berlin, Germany

^c Chair for Sociology, esp. Analysis of Social Structures, Otto-Friedrich-Universität Bamberg, Feldkirchenstraße 21, 96045, Bamberg, Germany

^d Mannheim Centre for European Social Research (MZES), Universität Mannheim, 68131, Mannheim, Germany

ARTICLE INFO

Keywords:

Self-fulfilling prophecy
Teacher expectations
Teacher feedback
Teacher-student interaction
Videography

ABSTRACT

This study investigated the link between teacher expectations and student learning, relying on longitudinal data from 64 classrooms and 1026 first-grade students in Germany. Further, based on a subsample of 19 classrooms with 354 students, we explored the mediating role of three characteristics of teacher feedback rated in video-recorded school lessons. The results showed that teacher expectations were inaccurate to some extent; that is, they did not entirely agree with students' current achievement, general cognitive abilities and motivations. In addition, this inaccuracy in teacher expectations significantly predicted students' end-of-year achievement, even after prior achievement, general cognitive abilities, motivation, and student background characteristics were considered. Specifically, inaccurately high teacher expectations were associated with greater achievement in reading and mathematics, whereas inaccurately low teacher expectations were associated with lower achievement in reading only. Furthermore, teacher feedback varied significantly with inaccurate teacher expectations but did not substantially mediate teacher expectancy effects.

1. Introduction

From the first day of school enrollment onward, teachers play a significant role in student learning (e.g., Dietrich, Dicke, Kracke, & Noack, 2015; Hattie, 2009; Roorda, Koomen, Spilt, & Oort, 2011). To promote student learning, it is crucial that teachers are aware of students' achievement, as well as their individual learning resources, as this knowledge is the basis for effective instructional decisions and enables teachers to provide sufficient support to individual students (e.g., Baumert & Kunter, 2013; Vogt & Rogalla, 2009). Such evaluations include not only perceptions of current student achievement but also expectations about students' learning and future achievement (Funder, 1995; Jussim, Robustelli, & Cain, 2009). At the same time, teachers' inaccurate achievement expectations can result in a self-fulfilling prophecy; that is, low expectations can hamper students' learning, whereas high expectations can foster students' learning and eventually lead to higher achievement gains. Rosenthal and Jacobson (1968), with their experiment "Pygmalion in the Classroom", were the first to provide evidence of a self-fulfilling prophecy (Merton, 1948) in the context of school. The experiment was controversial, with several scholars questioning different aspects of its results (Pellegrini & Hicks, 1972; Snow,

1969; Thorndike, 1968). Today, researchers generally agree that teacher expectancy effects exist (for overviews, see Jussim et al., 2009; Wang, Rubie-Davies, & Meissel, 2018). Teacher achievement expectations can thus affect child development from the first school days onward, as well as affecting later educational achievements and eventual outcomes.

Nonetheless, the question of how teacher expectancy effects emerge has not been conclusively answered. Most authors agree with the assumption that self-fulfilling prophecies in the classroom follow a sequence of three major steps (Jussim et al., 2009): (1) teachers form inaccurate expectations; (2) these expectations lead teachers to treat higher- and lower-expectancy students differently; and (3) students react to this differential teacher treatment in such a manner that confirms the initial teacher expectations, hence resulting in greater achievement gains for higher-expectancy students and lower achievement gains for lower-expectancy students. Until now, however, very few studies have provided empirical evidence of this three-step formation process; specifically, for the mediation of teacher expectancy effects by differential teacher behavior. Past studies have examined only one or two of the steps. Whereas there has been much research on accuracy in teacher expectations (for an overview, see Jussim et al.,

* Corresponding author.

E-mail address: sarah.gentrup@hu-berlin.de (S. Gentrup).

<https://doi.org/10.1016/j.learninstruc.2019.101296>

Received 13 March 2019; Received in revised form 23 November 2019; Accepted 25 November 2019

Available online 17 January 2020

0959-4752/ © 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

2009; for a meta-analysis, see Südkamp, Kaiser, & Möller, 2012) and the effects of teacher expectation inaccuracy on student achievement (for overviews, see Jussim, Eccles, & Madon, 1996; Jussim et al., 2009), relatively few studies have addressed the ways in which teachers communicate their expectations through their behavior, and even fewer studies have investigated whether such differential teacher behavior actually mediates teacher expectancy effects. Thus, there is a dearth of research on the effects of differential teacher behaviors triggered by teacher expectations on student outcomes. In the early years of expectancy research, four dimensions of teacher behavior were discussed as mediators of teacher expectancy effects: input (e.g., amount and difficulty of learning material provided), output (e.g., calling on a student), feedback (e.g., valence and elaborateness of feedback given) and climate (e.g., warmth and respectfulness in teacher-student interaction) (Rosenthal, 1973). Although empirical research has provided support for the relevance of each of these four dimensions (Harris & Rosenthal, 1985), evidence regarding the feedback dimension appears to be inconclusive. Findings on the relationships between different indicators of teacher feedback and teacher expectations, on the one hand, and student learning, on the other hand, have varied significantly, resulting in contradictory conclusions (Harris & Rosenthal, 1985). However, since there have been few newer empirical studies on the mediation of teacher expectancy effects, the recent research still relies on the four-factor model.

The present paper was based on a unique longitudinal data set that enabled us to examine the relationships between initial teacher expectations and later teacher feedback, as well as subsequent student achievement immediately at the beginning of students' school careers in Germany. Our study had two main goals. (1) By empirically tracing all three steps of the process potentially leading to the emergence of self-fulfilling prophecies in school, we sought to clarify whether associations between teacher expectations (as measured at the beginning of first grade) and student achievement (as measured at the end of first grade) are mediated by differences in teacher behavior, specifically teacher feedback. Thus, we directly investigated whether the three-step process happens as assumed. (2) Further, our study aimed to contribute to resolving open questions regarding the role of teacher feedback as a mediator of teacher expectancy effects. The inconclusive results in Harris and Rosenthal's meta-analysis show that this dimension of teacher behavior requires further investigation.

2. Theory and evidence

2.1. Inaccurate teacher expectations

Teacher expectations can be viewed as predictions of future student achievement (e.g., Jussim, 1986; Ready & Chu, 2015). Many researchers have only considered students' current achievements to be accurate bases of teacher expectations (e.g., Hinnant, O'Brien, & Ghazarian, 2009; McKown & Weinstein, 2002; Ready & Chu, 2015; Rubie-Davies & Peterson, 2016). From a pedagogical point of view, however, further student characteristics, such as general cognitive ability and motivation (in contrast to other student characteristics, such as gender or socioeconomic status, which should be irrelevant in a meritocratic school system) can also serve as valid predictors of teacher expectations (cf. de Boer, Bosker, & van der Werf, 2010). Based on these considerations and in line with research by Becker (2013) and de Boer et al. (2010), we characterize teacher expectations as accurate if they concur with actual students' achievement, general cognitive abilities and achievement motivation. Deviations in teacher expectations that exist beyond these learning-related student characteristics are hence referred to as teacher expectation inaccuracy.

Prior research has shown that teacher expectations for students' achievements are accurate to a substantial degree (Jussim et al., 2009). In meta-analyses, the shared variance between teacher judgments of current student achievement and students' actual achievements

amounted to approximately 40% (Hoge & Coladarci, 1989; Südkamp et al., 2012). The explained variance in teacher expectations was somewhat higher when general cognitive abilities and motivation were considered as valid predictors of teacher expectations (Becker, 2013; de Boer et al., 2010). Nevertheless, a significant portion of the variance in teacher expectations remained unexplained by students' learning-related characteristics. This variance reflects inaccuracy and has the potential to initiate self-fulfilling prophecies.

Furthermore, studies have shown that inaccuracy in teacher expectations does not occur randomly but is systematic for different groups of students. For example, negative bias in teacher expectations has been found for students from socially disadvantaged families (e.g., de Boer et al., 2010; Lorenz, Gentrup, Kristen, Stanat, & Kogan, 2016; Ready & Wright, 2011; Timmermans, de Boer, & van der Werf, 2016; Timmermans, Kuyper, & van der Werf, 2015; van Matre, Valentine, & Cooper, 2000), for ethnic minority students (e.g., Holder & Kessels, 2017; Lorenz et al., 2016; McKown & Weinstein, 2008; Meissel, Meyer, Yao, & Rubie-Davies, 2017; Morris, 2005; Ready & Wright, 2011), for boys and girls in gender-untypical domains (e.g., Hinnant et al., 2009; Holder & Kessels, 2017; Lorenz et al., 2016; Meissel et al., 2017; Ready & Wright, 2011; Riegler-Crumb & Humphries, 2012; Rubie-Davies & Peterson, 2016) as well as for students with special educational needs or learning disability statuses (e.g., Hurwitz, Elliott, & Braden, 2007; Jenkins & Demaray, 2016; Shifrer, 2013, 2016). Teacher expectancy effects based on such systematically biased expectations have the potential to contribute to educational inequalities (e.g., de Boer et al., 2010; Gentrup & Rjosk, 2018; Muntoni & Retelsdorf, 2018). However, as the main focus of the current study is the mediation of teacher expectancy effects via teacher feedback, we do not investigate group-specific bias but rather general teacher expectation inaccuracy.

2.2. Teacher expectations and differential teacher behavior

Most studies examining teacher behavior in the context of teacher expectancy effects are from the 1970s and 1980s (for a meta-analysis, see Harris & Rosenthal, 1985; for a review, see Jussim, 1986). In the last 30 years, very few studies investigated this issue (e.g., Chen, Thompson, Kromrey, & Chang, 2011; Ready & Chu, 2015; for a recent review, see Wang et al., 2018). Therefore, the contemporary expectancy research still relies on the four-factor model proposed by Rosenthal (1973), which was empirically evaluated in the meta-analysis by Harris and Rosenthal (1985). This model assumes that teachers' expectations affect four dimensions of their behavior (cf. Jussim et al., 2009). (1) Teachers might differ in the *input* they provide. For example, teachers might explain issues in a less complex manner to lower-expectancy students than to higher-expectancy students. (2) According to their expectations, teachers might provide different opportunities for students to produce *output*. For example, teachers might call on lower-expectancy students less frequently than they call on higher-expectancy students. (3) Teacher expectations might influence teacher *feedback*, which could be less positive and less constructive for lower-expectancy students than for higher-expectancy students. (4) Finally, the *climate* of teacher-student interactions might be less warm and respectful for lower-expectancy students than for higher-expectancy students. The results of the meta-analysis (Harris & Rosenthal, 1985) supported the relevance of each of the four proposed dimensions, with *climate* and *input* showing the strongest relationships with both teacher expectations and student outcomes. The evidence for *feedback*, however, remained inconclusive. The relationships of diverse indicators of teacher feedback with teacher expectations and student achievement varied significantly and showed, in part, contradicting results. The association between teacher expectations and *feedback* varied, depending on the indicators used, between $r = -.05$ and $r = 0.36$, showing in summary a small effect size ($r = 0.13$). The observed relationship between indicators of *feedback* and student achievement even ranged from $r = -.23$ to $r = 0.12$. Overall, the link between feedback and student achievement amounted

to $r = 0.07$. Thus, it appears that diverse indicators of feedback, collapsed into one feedback dimension, operate in opposite directions – with the consequence of an average effect size of approximately zero.

Despite providing important insights into the relationship between teacher expectations and teacher behavior, the meta-analysis also had its limitations. First, the authors did not ensure that the studies included in their meta-analysis controlled for actual student achievement. Therefore, it was not possible to determine whether the differential teacher treatment resulted from accurate teacher expectations or from teacher expectation inaccuracy. Second, only a few of the studies included in the meta-analysis investigated both the association between teacher expectations and teacher behavior and the association between teacher behavior and student outcomes. The meta-analysis thus did not directly address mediation effects (cf. Chow, 1987).

Newer studies on the relationships between teacher expectations and teacher behavior are few, and they largely suffer from the same limitations. Chen et al. (2011) investigated teacher expectancy effects on student self-concept in elementary schools in Taiwan and considered students' perceptions of oral feedback given by their teachers. The results showed that lower-expectancy students perceived receiving less positive and more negative oral feedback from their teachers than higher-expectancy classmates. Furthermore, the perceived feedback was significantly related to students' self-concept. However, this study also did not account for actual student achievement. Furthermore, it leaves open whether teacher feedback actually mediates the link between teacher expectations and students' self-concept.

Other newer studies examined teacher behavior in relation to teacher expectations at the classroom level (e.g., Rubie-Davies, Hattie, Townsend, & Hamilton, 2007; Rubie-Davies & Peterson, 2011). Although these studies supported the concept that teachers with high expectations for all of their students are more effective in teaching (e.g., provide more feedback and use more higher-order questions; Rubie-Davies, 2007), they also left open the question of whether the observed differences in teacher behavior actually mediate teacher expectancy effects.

Two newer studies directly investigated the mediation of teacher expectancy effects by teacher behavior. First, Urhahne (2015) analyzed the effects of teacher expectations on students' motivations and emotions. Differential teacher behaviors measured through student perceptions were found to partially mediate the link between teacher expectations and students' motivations and emotions. Second, Ready and Chu (2015) examined whether teachers' practice of ability grouping (which can be viewed as an indicator of *input*) mediates the link between individual teacher expectations and student achievement. Contradicting the findings by Harris and Rosenthal (1985) on the input dimension of teacher behaviors, ability grouping was found to be only a weak mediator of teacher expectancy effects (Ready & Chu, 2015). This limited and inconclusive evidence on the mediation of teacher expectancy effects via differential teacher behavior calls for research based on data that are better suited to addressing the theoretical arguments. The current study focuses on teacher feedback as a possible mediator of teacher expectancy effects.

2.2.1. Teacher expectations and teacher feedback

Teachers' expectations may shape their feedback practice on the basis of mental schemata. Based on their teaching experience, teachers can have internalized schemata of *high-achieving* and *low-achieving students*, that is, mental representations about typical characteristics and behaviors of high-achieving and low-achieving students, as well as perceptions about the appropriate teaching of these students (Fazio & Olson, 2014; Pendry, 2015). As an interview study of approximately 300 student teachers revealed, several positive attitudes are associated with high-achieving students. Such students are seen as interested and motivated, as exerting effort for their learning, and as showing disciplined behavior. Low-achieving students, in contrast, are perceived as undisciplined, uninterested, unmotivated and unintelligent (Schuchardt

& Dunkake, 2014). Teachers can interpret student behavior and achievement based on these mental representations of high-achieving and low-achieving students, and such interpretations can result in predetermined (patterns of) reactions (Dijksterhuis & Bargh, 2001).

According to the results of the study by Schuchardt and Dunkake (2014), the schema of a low-achieving student is connected, for instance, to the assumption of an undisciplined child. Teachers might thus be more sensitive to the misbehavior of students whom they categorize as low-achievers (that is, low-expectancy students) and notice misbehavior more often for these students than for students whom they categorize as high-achievers (that is, high-expectancy students). Consequently, teachers might more often give children for whom they have lower expectations feedback related to their behavior in the classroom, instead of feedback related to their performance.

As the schema of a high-achieving student also includes favorable student characteristics in general (cf. Schuchardt & Dunkake, 2014), teachers might interpret the achievement output of high-expectancy students to be more positive than the similar outputs of low-expectancy students (*confirmation bias*; Nickerson, 1998). That is, the same student achievement might trigger more or less positive performance feedback depending on the schema activated. Accordingly, teachers might more often praise students for whom they have higher expectations, whereas negative performance feedback might be more frequent for lower-expectancy students, independent of the students' actual achievements. Eventually, the valence of performance feedback might be more positive for higher-expectancy students than for lower-expectancy students.

As studies relying on attribution theory have indicated (Johnson, Feigenbaum, & Weiby, 1964), teachers also feel stronger self-efficacy in teaching high-achieving students. Therefore, when interacting with high-expectancy students, teachers' performance feedback might be more elaborate. For example, negative performance feedback might be more often accompanied by further elaboration and tips on how to proceed for higher-expectancy students than for lower-expectancy students. Teachers might also explain in greater detail how students have misunderstood the learning material or which types of errors they have made.

2.2.2. Role of teacher feedback in student learning

Educational research has supported the importance of teacher feedback for student learning (e.g., Bohn, Roehrig, & Pressley, 2004; Dean, Hubbell, Pitler, & Stone, 2012). In fact, teacher feedback has been empirically identified as being among the most important instructional practices for improving student learning. In a meta-analysis, high-quality feedback had an average effect size on student achievement of $d = 0.73$ and was among the top ten investigated instructional practices (Hattie, 2009). Additionally, when children were asked about possible sources of information regarding their own levels of achievement in school, they primarily referred to teacher feedback (Weinstein, 1983).

Feedback provides useful information about how well a student is performing. Frequent and informative teacher feedback therefore helps students to overcome mistakes and improve skills. In contrast, if feedback is less frequently provided or less informative, students might not be aware that they have not fully mastered the material and that they must improve their skills (c.f. Jussim et al., 2009). The literature has further emphasized that individualized teaching styles, in particular, require high-quality feedback. Such feedback must be specific, task-oriented and related to students' learning goals; it must evaluate how well students are performing to reach those goals; and it must provide information to students on how to proceed (Hattie & Timperley, 2007). Therefore, feedback related to students' performance should promote students' learning in reading and mathematics better than feedback related to students' behavior in class. To strengthen students' perceived self-efficacy, a higher intraindividual ratio of positive and therefore affirmative performance feedback should be pedagogically beneficial (cf. Bandura, 1994). Additionally, especially in the case of negative

Table 1
Research questions and hypotheses of the current study.

Research questions	Hypotheses
(1) Are teacher expectations of students' achievement inaccurate and, if so, to what extent?	(1) The variance in teachers' expectations cannot be fully explained by actual student skills, cognitive abilities, and motivation.
(2) Are inaccurate teacher expectations reflected in teachers' feedback practices?	(2) Higher values on the variable of teacher expectation inaccuracy are associated with: (a) more performance-related feedback (compared to feedback on student behavior); (b) more positive performance feedback; and (c) greater elaborateness of negative performance feedback.
(3) Do inaccurate teacher expectations predict student learning in reading and mathematics beyond other student characteristics that affect students' learning progress?	(3) In reading and mathematics: (a) inaccurately high expectations are related to higher achievement gains; and (b) inaccurately low expectations are related to lower achievement gains over time.
(4) Do teachers' feedback practices mediate teacher expectancy effects?	(4) Teacher feedback partially mediates teacher expectancy effects.

performance feedback, it is important that teachers provide detailed information about learning goals and how to proceed in learning.

2.3. Teacher expectancy effects on students' achievement

The effects of teacher expectations on student achievement have been extensively studied (for overviews, see Jussim et al., 1996; Jussim & Harber, 2005; Jussim et al., 2009). The bulk of research comes from the U.S. (e.g., Gill & Reynolds, 1999; Hinnant et al., 2009; Ma, 2001; McKown & Weinstein, 2008; Ready & Chu, 2015; Rubie-Davies et al., 2014), and a few studies were conducted in Europe (e.g., Germany: Friedrich, Flunger, Nagengast, Jonkmann, & Trautwein, 2015; Lorenz, 2018; Stahl, 2007; Netherlands: de Boer et al., 2010; Jungbluth, 1993, 1994). Despite significant methodological differences, most studies have found support for teacher expectancy effects on student achievement (cf. Wang et al., 2018). The effects in field studies varied between $r = 0.10$ and $r = 0.20$; in terms of the effect size d , the average effect was $d = 0.30$ (Jussim et al., 2009).

However, most studies on teacher expectancy effects examined students at later stages of their educational trajectories, that is, at the later grade levels of primary school or secondary school (e.g., Archambault, Janosz, & Chouinard, 2012; de Boer & van der Werf, 2015; Friedrich et al., 2015; Peterson, Rubie-Davies, Osborne, & Sibley, 2016; Zhou & Urhahne, 2013). Expectancy research focusing on students' early school career is comparatively sparse. Recently, Schenke, Nguyen, Watts, Sarama, and Clements (2017) analyzed the link between class-level teacher expectations and 4-year-old students' mathematical achievement. Ready and Chu (2015) investigated whether students' literacy achievement in kindergarten was affected by teacher expectations. The results of both studies suggest that teacher expectancy effects can manifest in such early years. Furthermore, research supports that teacher expectancy effects are stronger in new situations such as after school enrollment or school transitions (cf. Jussim et al., 1996; Jussim et al., 2009). For example, in the study by Kuklinski and Weinstein (2001), teacher expectancy effects were stronger in first grade than in the later grades of elementary school. Similarly, for the period of five years after the transition into secondary school, de Boer et al. (2010) found that the effects of teacher expectations were strongest in the first year, whereas they decreased somewhat in the second year and remained stable afterwards.

A common criticism of studies of expectancy effects is that teacher expectations might have predicted student outcomes simply because they were accurate representations of students' prior achievements (Jussim & Harber, 2005). The main challenge of measuring expectancy effects is thus to disentangle the statistical associations of student outcomes with accurate teacher expectations, on the one hand, from their associations with inaccurate teacher expectations, on the other hand. Only the latter association indicates teacher expectancy effects (de Boer et al., 2010). As Wang et al. (2018) outlined, approximately 40% of the studies on teacher expectancy effects conducted in the last 30 years did not consider actual student characteristics in their analyses and thus did not adequately account for this challenge.

In addition, most existing studies measured teacher expectations not at the beginning of the teacher-student interaction but after months or even years of such interaction (e.g., Friedrich et al., 2015; Hinnant et al., 2009; Peterson et al., 2016; Ready & Chu, 2015). In this situation, it is not possible to disentangle whether teacher expectations concur with actual student characteristics because they are accurate evaluations of these characteristics or because of preceding self-fulfilling prophecies. Against this background and the observation that approximately 40% of the studies did not account for actual student achievement (Wang et al., 2018), the validity of the existing results might be limited.

3. The current study

In summary, although evidence for the existence of teacher expectancy effects on student achievement abounds, only a few studies have sought to clarify how these effects emerge. Accordingly, there has been little research on teacher behaviors mediating the link between inaccurate teacher expectations and student achievement. In particular, the results for teacher behavior in terms of feedback have been inconclusive. Furthermore, methodological limitations might reduce the validity of research on expectancy effects. These limitations include a failure to consider baseline student achievements and the problem of measuring teacher expectations after prolonged periods of teacher-student contact.

With the aim of overcoming these shortcomings, the present study examined the associations of teacher expectations with teacher feedback and student learning, covering all three steps of the process potentially leading to the emergence of a self-fulfilling prophecy. We used data from a unique longitudinal study in Germany explicitly designed to investigate self-fulfilling prophecies in schools. Data collection started immediately after the beginning of the first school year, when preceding teacher-student interaction was minimal. We analyzed the processes underlying teacher expectancy effects in two domains: language and mathematics. More specifically, we studied teachers' feedback practices and explored whether they differed depending on teachers' expectations. We also investigated the extent to which feedback mediates teacher expectancy effects on student achievement in reading and mathematics. Table 1 contains the specific research questions that we addressed, as well as the corresponding hypotheses.

4. Methods

4.1. Databases

The study was based on data from the research project *Kompetenzerwerb und Lernvoraussetzungen*¹ (KuL; English translation:

¹ This interdisciplinary research project was conducted under the leadership of Prof. Dr. Cornelia Kristen (Otto-Friedrich-Universität Bamberg), Prof. Dr. Irena Kogan (Universität Mannheim), and Prof. Dr. Petra Stanat (Humboldt-Universität zu Berlin).

Competence Acquisition and Learning Preconditions). In the 2013/2014 school year, data were collected in 39 primary schools in North Rhine-Westphalia, Germany. The total sample of the research project included 1065 first graders from 67 classes (Kristen et al., 2018a). We excluded from the sample classes in which the teachers changed during the school year, which left $N = 1026$ students from $N = 64$ classes with $N = 67$ class and subject teachers in $N = 38$ schools. This sample is referred to as the *main sample*. In the main sample, on average, 16 students per class participated in the study ($SD = 5.25$), which corresponds to a participation rate of 68% ($SD = 21\%$).

The teachers (94% female) were, on average, 42 years old ($SD = 8.80$) and had an average work experience of twelve years ($SD = 8.89$). The teachers were predominantly nonimmigrant, meaning that they and both of their parents were born in Germany (94%). At the time of school enrollment, the participating students (48% female) in the main sample were, on average, 6 years and 5 months old ($SD = 0.33$). Based on information from the parent questionnaire, 36% of the children came from immigrant families (at least one parent born in a country other than Germany). The average families' socioeconomic background, as indicated by the Highest International Socio-Economic Index of Occupational Status (HISEI; Ganzeboom, 2010), was $M = 52.51$ ($SD = 19.44$).

A subsample of $n = 354$ children from $n = 19$ classrooms in $n = 13$ schools participated in the optional video study, which was conducted in the middle of the school year (Kristen et al., 2018b). Despite the voluntary participation of teachers and students, the subsample of classrooms did not differ significantly from the main sample in important characteristics. The teachers involved in the video study did not differ substantially from the main sample in terms of gender (95% female), immigrant status (95% nonimmigrant), age ($M = 41.85$, $SD = 9.51$) or years of professional experience ($M = 12.11$, $SD = 8.60$). The students in the video subsample (47% female) were also similar to those in the main sample regarding their age ($M = 6.48$, $SD = 0.35$) and socioeconomic background ($M = 52.03$, $SD = 19.69$). Slightly more students in the video subsample than in the main sample came from immigrant families (38%). In addition, the participation rate per class was substantially higher in the subsample ($M = 86\%$, $SD = 9\%$). Descriptive statistics of further student variables in the main sample and subsample are displayed in Appendix A.

4.2. Survey design

The KuL study was a mini-panel that tracked first graders and their teachers in the school subjects of German language and mathematics throughout the first grade. The study received approval from the research ethics committee of *Universität Mannheim*. Data collection was performed in three waves. The first wave occurred immediately after the beginning of the school year and included standardized achievement tests and interviews with the students (September to November 2013, during the third to ninth school weeks), a questionnaire for teachers (dispatched by the research team in September to November 2013, during the third to seventh school weeks) and telephone interviews with the parents (conducted in October to December 2013). The second survey wave in the middle of the school year (February and March 2014) included video recordings of teacher-student interactions during lessons and further interviews with the students. At the end of the school year, the students were tested and interviewed again (May and June 2014), and the teachers completed another questionnaire (dispatched by the research team at the beginning of May 2014). All three study waves and the instruments were pretested in two separate schools in the year preceding the main study.

The resulting data have two important advantages. First, both teacher expectations and student achievement were measured immediately after school enrollment. Hence, measured student achievement should not yet have been influenced by teachers' expectations and behaviors, as the teacher-student interaction prior to the data collection

was minimal. Second, the video recordings of classroom interactions allowed us to examine teacher feedback practices and link them to the teachers' expectations, on the one hand, and to student achievement, on the other hand. The data set thus enabled us to analyze the mediating mechanisms of self-fulfilling prophecies in schools considering a range of indicators of both teacher expectations and student characteristics.

4.3. Instruments

4.3.1. Teacher expectations

Teachers rated each of the participating students in their classes on five items (5-point-scale), indicating their expectations for each child's achievement in the German language (three items; $\alpha = 0.94$) and mathematics (two items; $\alpha = 0.94$). Teachers were asked to compare the skill levels that they expected the children to acquire by the end of first grade to that of their classmates (e.g., "Compared to his/her fellow students, how well do you expect this child to perform at the end of the school year? ... in German language/... in mathematics"; see Appendix B). Three of the items originated from the BiKS-3-10 study (Artelt, Blossfeld, Faust, Roßbach, & Weinert, 2013), and the other two were developed in the KuL study.

To investigate teacher expectation inaccuracy, we applied the residual approach proposed by Madon, Jussim, and Eccles (1997). The procedure is explained in section 4.4.2.

4.3.2. Student achievements and abilities

At the beginning and end of first grade, the students completed tasks of the subscales *phonological awareness* ($\alpha = 0.82$) and *reading* ($\alpha = 0.96$) from the *Fähigkeitsindikatoren Primarschule (FIPS)* computer-based assessment (German version of the *Performance Indicators in Primary Schools (PIPS)*; Bäuerlein et al., 2012) as measures of language skills and the subscale *mathematics* ($\alpha = 0.92$) as a measure of mathematical skills. Students also completed a deductive reasoning test (CFT; Weiß & Osterland, 1997; $\alpha = 0.78$) and the subscale *working memory* implemented in the FIPS assessment (Bäuerlein et al., 2012; $\alpha = 0.76$) at the beginning of the school year. The two scales captured students' general cognitive abilities.

4.3.3. Student motivation

In addition to students' achievement and general cognitive abilities, teachers might also rely on students' motivation when forming their expectations. In study wave 2, all participating children were interviewed regarding their enjoyment of learning (13 items; $\alpha = 0.78$) and the effort that they invest in learning (13 items; $\alpha = 0.70$). Both motivational traits were measured with an adapted version of a questionnaire by Rauer and Schuck (2004). For each item, the students indicated whether it applied to them (enjoyment of learning: e.g., "I like to learn at school"; effort: e.g., "I also try to solve very difficult tasks"). To increase differentiation, we used a 3-point scale (0 = *not true*, 1 = *partly true*, 2 = *completely true*) instead of the original dichotomous response scale (0 = *not true*, 1 = *true*). As indicated by mean scores greater than the central point of the response scales (see Appendix A), the first graders in this sample stated that they fully enjoyed learning and were exerting much effort in learning. This pattern is typical for young students (e.g., Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Spinath & Steinmayr, 2008).

4.3.4. Teacher feedback practice

In the middle of the school year, we video-recorded between two and four school lessons (approximately 45 min per lesson) in each class that participated in the video study. Usually, two lessons were recorded in each subject (i.e., German language and mathematics). We asked teachers to set up their lessons as usual. We also asked them to include phases in which they interact with the whole class as well as phases in which the students work on their own. As we scheduled the specific appointments for video-recording with each teacher, they knew

beforehand which lessons would be video-recorded. The videos were then coded by independent raters who had not been informed about any specific research questions or hypotheses of the study. The coding followed the method of event-sampling.

In a first step, the raters identified each time sequence in which the teacher directed his or her attention, an action or a statement toward one to three students at a time (coding of *interaction sequences*). Because the coding aimed at identifying differences in teacher behavior for individual students, time periods in which the teacher interacted with more than three students at a time (for example, with the whole class) were ignored. In a second step, each interaction sequence was chopped into subsequences (coding of *subsequences*) based on changes in the content of the interaction happening within the interaction sequence (e.g., feedback, elaborations, and instructions). In a final step, each subsequence was then rated in terms of its content (coding of *content*). Here, raters considered, i.e., a range of feedback codes (i.e., very positive/positive/negative/very negative performance/behavioral feedback) and elaboration codes (i.e., direct hint/prompt/supporting question).

The training of the raters was based on video recordings of school lessons from the pretest study. We conducted two training periods and assessed the raters' reliability after each period. For the first reliability check, each rater coded an exemplary video with predefined interaction sequences. Thus, the raters had to decide about the timing and content of the subsequences. The results were compared to a master coding performed by the first author of this study, who had also developed the coding guidelines. Estimates of the rater-master agreement are displayed in the upper half of [Appendix C](#). A second training period was conducted after the raters coded video recordings of 14 school lessons from the main study (21.5% of all video recordings). Part of this reliability check followed the same procedure as the first reliability check (i.e., video raters had to decide about the timing and content of subsequences within predefined interaction sequences). Additionally, the raters coded a second exemplary video in which not only the interaction sequences but also the timing of 168 subsequences were predefined. Here, the video raters had to focus exclusively on the content of the subsequences. Based on these two codings, we tested for three aspects of reliability, all indicating the average agreement of the raters' codings with the master coding (see the lower half of [Appendix C](#)). First, we split every interaction sequence coded by the video raters into 100-msec sections and calculated Cohen's kappa by comparing them with the same 100-msec sequences from the master coding. The resulting estimate indicates the content-related agreement between a rater's coding and the master coding, while considering differences in the coding of the timing of subsequences. Cohen's kappa varied between $\kappa = 0.62$ and $\kappa = 0.70$, indicating substantial agreement (see [Landis & Koch, 1977](#)). Second, we focused on the additional content-related rating of 168 subsequences that had been predefined with regard to their timing. The resulting kappa estimates amounted to values between $\kappa = 0.96$ and $\kappa = 0.97$, indicating high, almost perfect agreement. Finally, we investigated the correlations between the rater coding and master coding for each of the three teacher feedback variables used in the analyses (see the following paragraphs for more information). The correlations varied between $r = 0.70$ and $r = 0.97$. These scores indicate medium to high agreement.

The coding of the videos focused on different contents of the interactions that occurred between teachers and students. One type of content relevant for the present study was feedback, which was further categorized into performance feedback (e.g., "You read very well") and behavioral feedback (e.g., "Great! You really worked hard"). In addition, whether the feedback was very positive, positive, negative or very negative was coded. Because very positive and very negative feedback rarely occurred, we collapsed the categories into positive versus negative performance and behavioral feedback.

Based on the theoretical considerations (see sections 2.2.1 and 2.2.2), in the present study, we investigated three characteristics of teacher feedback: (1) the *performance relatedness of feedback*, that is, the extent to

which teacher feedback was performance-related and not related to the students' behavior; (2) the *valence of performance feedback* that a child received, that is, the extent to which the performance feedback was positive and not negative; and (3) the *elaborateness of negative performance feedback*, that is, whether in the case of negative performance feedback, the teacher provided the child with further information about how to proceed. To consider differences in the duration of school lessons and in the number of interactions per child, we calculated intraindividual percentages for these three characteristics of teacher feedback.

First, the *performance relatedness of feedback* was calculated via the ratio of performance feedback to total feedback that a child received. A value of 0.5 on this measure indicates that a child received the same amount of performance feedback as of behavioral feedback. Higher values indicate that a child received more performance feedback than behavioral feedback (e.g., a value of 0.8 means that the ratio of received feedback was 80% performance and 20% behavioral feedback).

Second, the *valence of performance feedback* was calculated via the ratio of positive performance feedback to total performance feedback for each child. Values greater than 0.5 on this measure indicate that a child received more positive and therefore affirmative performance feedback than negative performance feedback.

Third, the *elaborateness of negative performance feedback* was calculated via the ratio of negative performance feedback with further elaboration to total negative performance feedback. That is, for negative performance feedback (e.g., pointing to a mistake), we determined whether the child received at least one tip or suggestion from the teacher in the same interaction. This was the case if the code "very negative/negative performance feedback" and one of the elaboration codes (i.e., direct hint/prompt/supporting question) occurred in the same interaction sequence. A higher value on this measure indicates that negative performance feedback was typically accompanied by further elaboration, while lower values point to negative performance feedback with little or no advice for improvement.

4.3.5. Background variables

Students' background characteristics can affect learning progress in school, as well as teacher expectations. To account for these influences, we considered two aspects of students' social background: socioeconomic status and parental education. The *Highest International Socio-Economic Index of Occupational Status* among the parents (HISEI; [Ganzeboom, 2010](#)) was used to represent the socioeconomic status of the family. The HISEI is an internationally established measure ([Baumert & Maaz, 2006](#); [OECD, 2003](#)) that quantifies the attributes of occupations that convert education into income ([Ganzeboom, Graaf, & Treiman, 1992](#)). Parental education was captured by a dummy-coded variable differentiating between families with at least one parent with an *Abitur* (higher education entrance qualification in Germany), coded as 1, and parents without the *Abitur*, coded as 0.

We further controlled for whether the students came from immigrant families, defined as having at least one parent born in a country other than Germany. Immigrant status was included in the analyses as a dummy variable (0 = *nonimmigrant family*, 1 = *immigrant family*). Furthermore, students' gender (0 = *male*, 1 = *female*) and students' age in years at the start of school enrollment (grand-mean centered, that is, the average student age constitutes the zero point) served as control variables when investigating students' learning progress.

4.4. Analytic strategy

4.4.1. General information

For all of the analyses we used Stata/SE data analysis software, version 15.1 ([StataCorp LLC, 1985–2017](#)). Variables with missing information were imputed under the missing at random assumption using the fully conditional specification ([van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006](#)). The percentage of missing data varied between 3 and 7% for achievement-related student variables

(achievement, ability, motivation), between 2 and 6% for teacher expectations, between 13 and 17% for student background variables, and between 1 and 16% for teacher feedback. The imputation models included not only the variables used in the analyses but also further information from teacher and parent interviews. Missing data were imputed separately for the main sample and the video subsample. All of the descriptive and regression analyses were conducted individually for 50 imputed data sets, and their parameters were subsequently pooled according to Rubin's rules (Rubin, 1987). To calculate the standardized regression coefficients, we re-ran all models with z-standardized metric variables. In all of the regression analyses, we used the cluster command in Stata (StataCorp, 2017) to consider the clustered structure of the data (students within classrooms). This procedure produces cluster-robust standard errors (which are based on the *sandwich estimator of variance* developed by Huber, 1967 and White, 1980, 1982), which account for heteroskedasticity and dependencies within nested data.

4.4.2. Separating teacher expectation inaccuracy

As a first step, we explored teacher expectation inaccuracy (Research Question 1), which we defined as teacher expectations that deviate from students' prior achievement, general cognitive abilities and motivation. To identify such inaccuracy at the beginning of the school year, we applied the residual approach proposed by Madon et al. (1997). We conducted multiple regression analyses with cluster-robust standard errors predicting teacher expectations regarding German language and mathematics performance from students' achievements (German language: phonological awareness and reading achievement; mathematics: mathematical achievement), general cognitive abilities (deductive reasoning and working memory), and motivation (enjoyment of learning and effort). The residuals resulting from these regressions reflect the variance that remained unexplained and served as our measure of teacher expectation inaccuracy. Residual scores around zero indicated accurate expectations based on the aforementioned student characteristics, positive values indicated inaccurately high expectations, and negative values indicated inaccurately low expectations.

In addition to the metric residual score, we used, in separate analyses, a series of dummy variables distinguishing between *inaccurately high expectations* (residual score more than one standard deviation greater than the mean), *inaccurately low expectations* (residual score more than one standard deviation less than the mean) and *accurate expectations* (residual score within one standard deviation of the mean). In these analyses, *accurate expectations* served as the reference category. We used a tolerance criterion of one standard deviation, as it is the changing point in normal distributions. However, as the results may differ depending on the exact criterion chosen, we additionally ran all models with 0.5 standard deviations as the cut-off for robustness checks (results presented in footnotes).

4.4.3. Analyzing teacher expectancy effects

To estimate the relationship between teacher expectation inaccuracy and teacher feedback (Research Question 2), we performed linear regression analyses with cluster-robust standard errors for each type of teacher feedback using the subsample. The models controlled for students' achievement at the beginning of the school year, general cognitive abilities, and motivation.

Subsequently, to investigate teacher expectancy effects (Research Question 3), we ran linear regression models with cluster-robust standard errors based on the main sample separately for students' reading achievement and their mathematical achievement. In addition to students' prior achievements, general cognitive abilities and motivation, in these models, we controlled for students' age, gender and family background, as these variables are known to predict students' learning progress. Since teacher expectation inaccuracy correlates with some of these student characteristics, not controlling for them would have caused the statistical relations between the teacher expectation inaccuracy and *later* student achievement to be confounded by background-related influences that existed independently from teacher

expectations (e.g., advantageous home learning environments in families with a higher socioeconomic status). In the first model, we included the metric teacher expectation inaccuracy as a predictor. As higher scores indicated either inaccurately high expectations or less inaccurately low expectations, we expected a positive regression coefficient for this variable on achievement. The second model included the dummy variables for inaccurately high and low teacher expectations, with the accurate-expectations category as the reference. This model allowed us to examine differences in the effect sizes of inaccurately high and inaccurately low teacher expectations. We expected positive regression coefficients for inaccurately high expectations and negative regression coefficients for inaccurately low expectations.

Finally, we directly investigated the assumed mediation of teacher expectancy effects through teacher feedback (Research Question 4). We explored whether teacher feedback predicted students' achievement development while controlling for teacher expectation inaccuracy, as well as whether the direct link between teacher expectation inaccuracy and student achievement significantly decreased when considering teacher feedback. Fig. 1 illustrates our overall analytic strategy.

5. Results

5.1. Teacher expectations and teacher expectation inaccuracy

Table 2 displays descriptive statistics for teachers' expectations and teachers' expectation inaccuracy in the main sample and the subsample. On average, teachers' expectations were slightly above the middle of the scale (German language: $M = 3.31$, $SD = 0.97$, mathematics: $M = 3.36$, $SD = 0.88$). That is, teachers assumed that their students performed somewhat above the classroom average overall.

Regression analyses of teacher expectations based on the main sample revealed that approximately 35% of the variation in teacher expectations was explained by students' achievements, general cognitive abilities, and motivation (German language: 33.85%, mathematics: 37.08%; models not displayed)². This result indicates, in accordance with Hypothesis 1, that approximately 65% of the variation in teacher expectations remained unexplained and was interpreted as teacher expectation inaccuracy. However, it is possible that some part of this unexplained variance was due to teacher observations of student characteristics that were not measured in this study or to measurement error. We will further debate this issue in the discussion section. According to the descriptive results displayed in Table 2, the teacher expectation inaccuracy varied substantially between students in both domains, indicating that some teacher evaluations overlapped more than others with actual student achievement, ability and motivation. The intraclass correlations showed that the major part of the variation occurred within classes (German language: $ICC = 0.08$, mathematics: $ICC = 0.05$). Descriptive statistics of the categorical teacher expectation inaccuracy variable revealed that approximately 70% of teacher expectations differed not more than one standard deviation from zero and were thus, based on this tolerance criterion, defined as accurate (German language: 66.81%, mathematics: 70.80%). At the same time, approximately 15% of teacher expectations were categorized as either inaccurately low or inaccurately high (German language: 16.04% inaccurately low, 17.14% inaccurately high; mathematics: 14.62% inaccurately low, 14.58% inaccurately high).³

To ensure that the metric residual scores were unrelated to students'

² For further information on similar models, see Lorenz et al. (2016).

³ When the categorization was based on a tolerance criterion of 0.5 standard deviations (instead of one standard deviation), approximately 40% of teacher expectations were categorized as accurate (German language: 37.86%, mathematics: 39.99%). Correspondingly, approximately 30% were categorized as either inaccurately high (German language: 31.36%, mathematics: 30.35%) or inaccurately low (German language: 30.78%, mathematics: 29.66%).

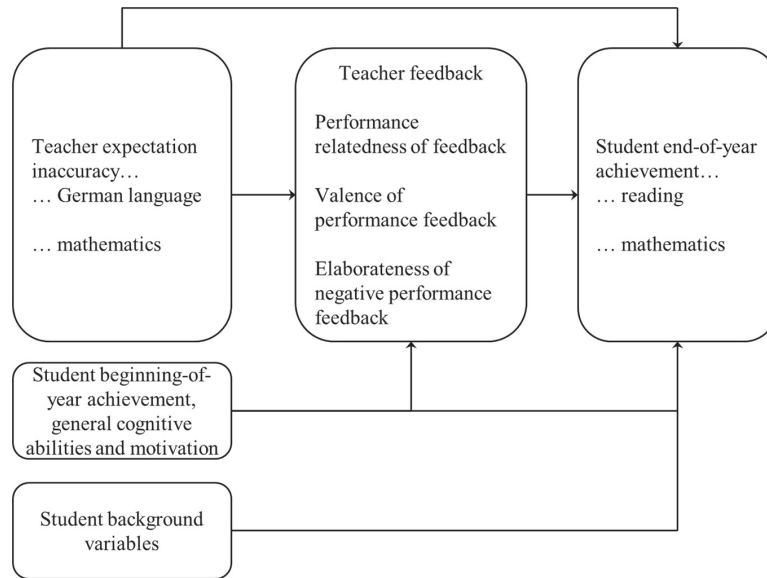


Fig. 1. Illustration of the overall analytic strategy.

initial achievements (as they should be, because initial achievement, general cognitive abilities, and motivation were controlled for in the regressions because we defined them as accurate influences on teacher expectations), we checked for bivariate correlations between the residual scores and students' beginning-of-year achievements. Except for a weak correlation in the subsample between the teacher expectation inaccuracy in the German language domain and phonological awareness ($r = 0.17$, $p = .002$), the correlations were all nonsignificant.

5.2. Teacher feedback

Based on the subsample, Table 2 also contains descriptive statistics of teacher feedback. During the videotaped lessons, students received, on average, individual feedback about their performance or their

behavior 23 times ($M = 23.02$, $SD = 18.97$). On average, 66.69% of the feedback that a student received was related to his/her performance, and 33.31% was related to his/her behavior. The ratio between performance and behavioral feedback varied substantially between students ($SD = 21.33$). Some students received only performance feedback ($max = 100.00\%$), and some students received only behavioral feedback ($min = 0.00\%$). Taking a closer look at the valence of performance feedback revealed that positive performance feedback was more prevalent than negative performance feedback. For a student who received at least one instance performance feedback during the videotaped lessons, the feedback was positive in 72.57% of the instances ($SD = 19.51$). Negative performance feedback accompanied by further elaboration (e.g., how to fix a mistake) occurred in 42.52% of the interactions that included at least one instance of negative performance

Table 2
Descriptive statistics of teacher expectations, teacher expectation inaccuracy and teacher feedback.

Variables	Main sample					Subsample				
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Teacher expectations										
German language	1026	3.31	0.97	1	5	354	3.31	1.00	1	5
Mathematics	1004	3.36	0.88	1	5	354	3.38	0.89	1	5
Teacher expectation inaccuracy										
German language	1026	0.00	0.79	−2.45	2.16	354	0.04	0.78	−2.45	2.15
Mathematics	1004	0.00	0.69	−2.47	2.58	354	0.04	0.70	−2.47	2.55
Categories of expectation inaccuracy	%					%				
German language										
Inaccurately low expectations	16.04					15.53				
Accurate expectations	66.81					66.11				
Inaccurately high expectations	17.14					18.37				
Mathematics										
Inaccurately low expectations	14.62					12.92				
Accurate expectations	70.80					69.60				
Inaccurately high expectations	14.58					17.49				
						<i>n</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Teacher feedback (FB) practice										
Absolute amount of FB						354	23.02	18.97	0	108
Performance relatedness of FB						351	66.69	21.33	0	100
Valence of performance FB						348	72.57	19.51	0	100
Elaborateness of negative performance FB						297	42.52	34.34	0	100

Table 3
Regression of three characteristics of teacher feedback (FB) on teacher expectation inaccuracy.

Variable	Performance relatedness of FB				Valence of performance FB				Elaborateness of negative performance FB			
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	<i>b</i> (SE)	β	<i>b</i> (SE)	β	<i>b</i> (SE)	β	<i>b</i> (SE)	β	<i>b</i> (SE)	β	<i>b</i> (SE)	β
Teacher expectation inaccuracy												
German language	3.39 (1.44)	.12*			2.58 (1.12)	.10*			-5.03 (2.43)	-.11*		
Mathematics			3.95 (1.84)	.13*			2.84 (1.11)	.10*			-3.58 (2.81)	-.07
Student achievement t0												
Phonological awareness	0.09 (0.23)	.02			0.40 (0.31)	.09			-0.27 (0.67)	-.04		
Reading	0.02 (0.06)	.03			0.13 (0.05)	.15*			0.02 (0.12)	.01		
Mathematics			-0.06 (0.20)	-.02			0.32 (0.17)	.14*			-0.43 (0.27)	-.11
Student general cognitive abilities												
Deductive reasoning	0.78 (0.45)	.11	0.92 (0.55)	.13	0.21 (0.38)	.03	0.30 (0.39)	.05	-1.05 (0.79)	-.09	-0.62 (0.87)	-.05
Working memory	0.79 (0.67)	.06	0.85 (0.64)	.07	0.13 (0.82)	.01	-0.07 (0.82)	-.01	-3.09 (1.08)	-.15*	-2.71 (1.01)	-.13*
Student motivation												
Enjoyment of learning	8.38 (4.31)	.15*	8.77 (4.09)	.15*	0.38 (3.40)	.01	0.97 (3.41)	.02	-8.16 (6.59)	-.09	-8.25 (6.54)	-.09
Effort	-2.64 (6.58)	-.03	-2.19 (6.56)	-.03	4.20 (4.09)	.06	4.63 (4.05)	.07	7.75 (9.21)	.06	8.98 (8.87)	.07
Intercept	49.08*** (8.05)		49.88*** (7.66)		55.05*** (6.05)		52.84*** (5.41)		61.20*** (9.33)		63.45*** (8.98)	
R-squared	6.26%		6.23%		7.93%		5.25%		5.66%		5.48%	
<i>n</i>	351		351		348		348		297		297	
Cluster	19		19		19		19		19		19	

Note. t0 = beginning of first grade. * $p < .10$. ** $p < .05$. *** $p < .001$.

feedback ($SD = 34.34$). Table 2 also reveals that the sample sizes for the three characteristics of teacher feedback varied. This variation is due to the fact that some students received no feedback ($n = 3$), no performance feedback ($n = 3$) or no negative performance feedback ($n = 51$) during the video-recorded school lessons. Thus, for these students, we could not calculate intraindividual percentages (see section 4.3.4 for more information) and had to exclude them from the respective analyses.

5.2.1. Teacher expectation inaccuracy and teacher feedback

Based on the subsample, we investigated the relation between teacher expectation inaccuracy and teacher feedback. The results of the multivariate regressions with cluster-robust standard errors displayed in Table 3 indicate positive relationships between teacher expectation inaccuracy and two of the three characteristics of teacher feedback, even after students' achievement, general cognitive abilities, and motivation were controlled. In accordance with Hypotheses 2a and 2b, higher values on the variable of teacher expectation inaccuracy (that is, inaccurately higher expectations or less inaccurately low expectations) were associated with a higher intraindividual ratio of performance feedback than of behavioral feedback (German language: $\beta = 0.12$, $p = .031$; mathematics: $\beta = 0.13$, $p = .047$) and with more positive than negative performance feedback (German language: $\beta = 0.10$, $p = .035$; mathematics: $\beta = 0.10$, $p = .021$). That is, if the value of the teacher expectation inaccuracy increased by 1 point, approximately 4% more of the feedback a student received was performance feedback and not behavioral feedback; similarly, the percentage of positive performance feedback compared to negative performance feedback increased by approximately 3% when the teacher expectation inaccuracy increased by 1 point. In other words, regardless of students' actual achievements, teachers tended to give more performance feedback (in comparison to behavioral feedback) and more positive than negative performance feedback to students whom they inaccurately expected to show higher achievement. Conversely, equally achieving lower-expectancy students received less performance feedback and more behavioral feedback from their teachers, and the performance feedback was less likely to be positive. Contradicting Hypothesis 2c, teachers did not provide more elaboration in the case of negative performance feedback for their inaccurately higher-expectancy students. On the contrary, the results tend to point in the opposite direction, indicating more elaboration for inaccurately lower-expectancy students. However, this weak association is significant on the 10%-level only in the German

language domain (German language: $\beta = -.11$, $p = .055$; mathematics: $\beta = -.07$, $p = .222$).

5.3. Teacher expectancy effects on student learning

Based on the main sample, we further examined the direct link between teacher expectation inaccuracy and student learning (see Table 4). In the German language domain, the metric residual expectancy score significantly predicted reading achievement at the end of the school year (model 1, $\beta = 0.21$, $p < .001$). Consistent with Hypothesis 3, this result indicates that higher scores on teacher expectation inaccuracy (that is, inaccurately higher expectations or less inaccurately low expectations) in the German language domain were associated with higher reading skills at the end of first grade, even after students' prior achievements, general cognitive abilities, motivation, gender, age, and family backgrounds were considered. The variation in the end-of-school-year reading achievement that was additionally explained by teacher expectation inaccuracy amounted to 4.07% (model without teacher expectation inaccuracy not displayed: $R^2 = 25.60\%$; model 1: $R^2 = 29.67\%$). We further investigated the regression coefficients of inaccurately high expectations and the coefficients of inaccurately low expectations, while accurate expectations served as the reference category. Model 2 confirmed that, compared to accurate expectations, inaccurately high expectations were related to higher end-of-year reading achievement ($\beta = 0.25$, $p = .002$), and inaccurately low expectations were related to lower ($\beta = -.42$, $p < .001$) end-of-year reading achievement.

In mathematics, the association between teacher expectation inaccuracy and student achievement was somewhat less pronounced. Teacher expectation inaccuracy significantly predicted gains in end-of-year mathematical achievement (model 3, $\beta = 0.09$, $p = .001$) when controlling for students' prior achievements, general cognitive abilities, motivation, gender, age, and family background (0.62% of the variance was additionally explained by the teacher expectation inaccuracy; model without teacher expectation inaccuracy not displayed: $R^2 = 57.13\%$; model 3: $R^2 = 57.75\%$). Moreover, comparing inaccurately high and inaccurately low expectations to accurate expectations revealed only small differences in students' achievements (model 4). Consistent with Hypothesis 3a, students who were exposed to inaccurately high expectations gained somewhat more mathematical skills during first grade ($\beta = 0.14$, $p = .014$). However, in contrast to Hypothesis 3b, students for whom their teachers had inaccurately low

Table 4
Teacher expectation inaccuracy as a predictor of student achievement.

Variable	Reading				Mathematics			
	Model 1		Model 2		Model 3		Model 4	
	<i>b</i> (<i>SE</i>)	β	<i>b</i> (<i>SE</i>)	β	<i>b</i> (<i>SE</i>)	β	<i>b</i> (<i>SE</i>)	β
Teacher expectation inaccuracy								
German language								
Metric inaccuracy	9.91 (1.79)	.21***						
Inaccurately high expectations			9.35 (2.81)	.25**				
Inaccurately low expectations			−15.93 (3.69)	−.42***				
Mathematics								
Metric inaccuracy					0.89 (0.25)	.09***		
Inaccurately high expectations							0.98 (0.39)	.14*
Inaccurately low expectations							−0.71 (0.53)	−.10
Student achievement t0								
Phonological awareness	1.93 (0.28)	.24***	1.87 (0.28)	.23***				
Reading	0.05 (0.04)	.03	0.04 (0.04)	.03				
Mathematics					0.54 (0.03)	.62***	0.54 (0.03)	.62***
Student general cognitive abilities								
Deductive reasoning	2.05 (0.41)	.16***	2.04 (0.41)	.16***	0.28 (0.06)	.11***	0.28 (0.06)	.11***
Working memory	1.26 (0.55)	.06*	1.24 (0.57)	.06*	0.06 (0.09)	.02	0.07 (0.09)	.02
Student Motivation								
Enjoyment of learning	3.28 (4.30)	.03	3.26 (4.47)	.03	0.69 (0.60)	.04	0.71 (0.61)	.04
Effort	23.75 (5.99)	.18***	23.56 (6.03)	.18***	0.96 (0.90)	.04	0.86 (0.91)	.03
Student background								
Student age	−0.54 (3.66)	−.00	−0.53 (3.67)	−.00	−0.55 (0.51)	−.03	−0.52 (0.51)	−.02
Female gender	0.86 (1.90)	.02	1.40 (1.90)	.04	−1.79 (0.32)	−.25***	−1.79 (0.32)	−.26***
HISEI	0.16 (0.07)	.08*	0.17 (0.07)	.09*	0.01 (0.01)	.04	0.01 (0.01)	.04
Abitur	0.49 (2.68)	.01	1.00 (2.69)	.03	0.36 (0.36)	.05	0.43 (0.36)	.06
Immigrant status	−0.21 (2.78)	−.01	−0.19 (2.70)	−.00	−0.17 (0.36)	−.02	−0.14 (0.36)	−.02
Intercept	9.16 (9.01)		10.65 (9.08)		19.46*** (1.23)		19.43*** (1.22)	
R-squared	29.67%		29.27%		57.75%		57.43%	
N	1026		1026		1004		1004	
Cluster	64		64		63		63	

Note. t0 = beginning of first grade. ⁺ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

expectations did not differ from their accurately estimated classmates ($\beta = -.10$, $p = .188$).⁴

5.4. Mediating role of teacher feedback

To investigate the mediating role of teacher feedback in the expectancy effects identified above, we first examined whether teacher feedback predicted student achievement in reading and mathematics at the end of the school year after controlling for teacher expectation inaccuracy, as well as students' prior achievements, general cognitive abilities, motivation, gender, age, and family background (see Tables 5 and 6). The multivariate regression analyses with cluster-robust standard errors were based on the subsample. The analyses revealed a positive association between the valence of performance feedback and mathematical achievement at the end of the school year ($\beta = 0.08$, $SE = 0.03$, $p = .020$, 95% CI [0.01, 0.14]); that is, the more positive and the less negative performance feedback a student received, the more mathematical skills the student gained during first grade. In reading, the valence of performance feedback did not predict students' learning ($\beta = 0.04$, $SE = 0.04$, $p = .382$, 95% CI [−0.05, 0.13]). This result was also observed for the other two indicators of teacher feedback: no significant regression coefficients were found for the

performance relatedness of feedback (reading: $\beta = 0.07$, $SE = 0.07$, $p = .307$, 95% CI [−0.08, 0.22]; mathematics: $\beta = 0.02$, $SE = 0.04$, $p = .547$, 95% CI [−0.06, 0.11]) and the elaborateness of negative performance feedback (reading: $\beta = -.02$, $SE = 0.06$, $p = .730$, 95% CI [−0.15, 0.11]; mathematics: $\beta = 0.02$, $SE = 0.04$, $p = .607$, 95% CI [−0.07, 0.11]) for reading or mathematical achievement.

Further, we investigated the potential mediation of the teacher expectancy effect in the German language (Table 5) and the mathematical domain (Table 6) by the three characteristics of teacher feedback. We compared the coefficient of teacher expectation inaccuracy in model 0, which shows the expectancy effect in the subsample, to the coefficients of teacher expectation inaccuracy in models 1, 2 and 3 in the same table, each of which additionally controlled for teacher feedback. For all three teacher feedback variables, the results did not indicate a substantial mediation of teacher expectancy effects, which contradicted Hypothesis 4 (for all changes in the coefficient of teacher expectation inaccuracy, $p > .10$). For example, the direct effect of inaccurate teacher expectations on mathematical end-of-year achievement decreased marginally when additionally controlling for the valence of performance feedback, from $b = 0.97$, $\beta = 0.10$, $SE = 0.04$, $p = .030$, 95% CI [0.01, 0.18] to $b = 0.93$, $\beta = 0.09$, $SE = 0.04$, $p = .041$, 95% CI [0.00, 0.18]. The reduction was not statistically significant.

Based on considerations by Zhao, Lynch, and Chen (2010), we additionally examined the indirect effects of teacher expectation inaccuracy on student achievement via each of the three teacher feedback variables in order to examine mediation. We calculated the indirect effects with structural equation models using a full information maximum likelihood procedure. Since the standard errors of indirect effects are non-normally distributed (Jose, 2013), we applied a Monte Carlo approach using the Stata package “medsem” (Mehmetoglu, 2018) in

⁴ When the categorization of the teacher expectation inaccuracy was based on a tolerance criterion of 0.5 standard deviations (instead of one standard deviation), the results point in the same direction, although the effect sizes were somewhat lower (German language: inaccurately high expectation: $\beta = 0.17$, $p = .039$, inaccurately low expectation: $\beta = -.32$, $p < .001$; mathematics: inaccurately high expectation: $\beta = 0.14$, $p = .005$, inaccurately low expectation: $\beta = -.08$, $p = .188$).

Table 5

Mediation of the relationship between teacher expectation inaccuracy and students' reading achievement.

	Model 0		Model 1		Model 2		Model 3	
Variable	<i>b</i> (SE)	β	<i>b</i> (SE)	β	<i>b</i> (SE)	β	<i>b</i> (SE)	β
Teacher expectation inaccuracy								
German language	8.50 (3.25)	.19*	8.37 (3.32)	.18*	8.63 (3.39)	.19*	9.20 (3.30)	.20*
Teacher feedback (FB)								
Performance relatedness of FB			0.13 (0.12)	.07				
Valence of performance FB					0.07 (0.08)	.04		
Elaborateness of negative performance FB							−0.02 (0.06)	−.02
Student achievement t0								
Phonological awareness	2.36 (0.53)	.29***	2.41 (0.53)	.30***	2.38 (0.52)	.30***	2.32 (0.48)	.29***
Reading	−0.03 (0.07)	−.02	−0.04 (0.07)	−.02	−0.04 (0.07)	−.03	−0.01 (0.08)	−.01
Student general cognitive ability								
Deductive reasoning	1.10 (0.56)	.09*	1.07 (0.55)	.09*	1.23 (0.57)	.10*	1.28 (0.67)	.11*
Working memory	2.58 (0.92)	.12*	2.48 (0.91)	.11*	2.51 (0.92)	.12*	2.11 (1.02)	.10*
Student motivation								
Enjoyment of learning	−6.74 (6.94)	−.07	−6.27 (7.02)	−.07	−5.50 (6.87)	−.06	−4.42 (8.56)	−.05
Effort	30.58 (10.04)	.24**	27.76 (10.09)	.22*	27.05 (9.96)	.21*	29.24 (10.80)	.23*
Student background								
Student age	4.69 (7.55)	.05	4.39 (7.86)	.04	4.70 (7.98)	.05	6.52 (8.18)	.06
Female gender	3.62 (3.17)	.10	2.21 (3.56)	.06	3.54 (3.46)	.10	3.02 (3.39)	.08
HISEI	0.15 (0.12)	.08	0.17 (0.11)	.10	0.17 (0.11)	.09	0.13 (0.12)	.07
Abitur	−3.18 (4.52)	−.09	−3.64 (4.38)	−.10	−2.85 (4.71)	−.08	−1.19 (5.09)	−.03
Immigrant status	2.06 (3.85)	.06	2.48 (3.98)	.07	2.40 (4.01)	.07	2.63 (3.96)	.07
Intercept	13.02 (14.54)		8.30 (13.54)		10.15 (15.31)		12.15 (14.26)	
R-squared	31.17%		31.43%		31.42%		31.76%	
<i>n</i>	354		351		348		297	
Cluster	19		19		19		19	

Note. t0 = beginning of first grade. ⁺*p* < .10. * *p* < .05. ** *p* < .01. *** *p* < .001.**Table 6**

Mediation of the relationship between teacher expectation inaccuracy and students' mathematical achievement.

	Model 0		Model 1		Model 2		Model 3	
Variable	<i>b</i> (SE)	β	<i>b</i> (SE)	β	<i>b</i> (SE)	β	<i>b</i> (SE)	β
Teacher expectation inaccuracy								
Mathematics	0.97 (0.40)	.10*	0.97 (0.43)	.10*	0.93 (0.42)	.09*	0.93 (0.43)	.09*
Teacher feedback (FB)								
Performance relatedness of FB			0.01 (0.01)	.02				
Valence of performance FB					0.03 (0.01)	.08*		
Elaborateness of negative performance FB							0.00 (0.01)	.02
Student achievement t0								
Mathematics	0.54 (0.05)	.66***	0.54 (0.05)	.66***	0.53 (0.05)	.64***	0.55 (0.05)	.67***
Student general cognitive abilities								
Deductive reasoning	0.20 (0.09)	.08*	0.20 (0.09)	.09*	0.21 (0.09)	.09*	0.25 (0.10)	.11*
Working memory	0.10 (0.17)	.02	0.09 (0.17)	.02	0.09 (0.17)	.02	0.14 (0.18)	.03
Student motivation								
Enjoyment of learning	0.07 (0.87)	.00	0.09 (0.88)	.00	0.17 (0.87)	.01	−0.13 (0.93)	−.01
Effort	1.22 (1.61)	.05	1.14 (1.67)	.05	1.02 (1.66)	.04	1.22 (1.66)	.05
Student background								
Student age	−0.53 (0.60)	−.03	−0.56 (0.61)	−.03	−0.57 (0.63)	−.03	−0.20 (0.69)	−.01
Female gender	−1.06 (0.49)	−.15*	−1.17 (0.51)	−.17*	−1.25 (0.51)	−.18*	−0.96 (0.49)	−.14*
HISEI	0.03 (0.02)	.10*	0.03 (0.02)	.09*	0.03 (0.02)	.09	0.03 (0.02)	.09*
Abitur	−0.64 (0.53)	−.09	−0.68 (0.53)	−.10	−0.62 (0.54)	−.09	−0.63 (0.57)	−.09
Immigrant status	0.08 (0.64)	.01	0.09 (0.65)	.01	−0.01 (0.64)	−.00	0.19 (0.69)	.03
Intercept	19.48*** (2.31)		19.29*** (2.33)		18.41*** (2.45)		18.73*** (2.50)	
R-squared	59.69%		59.40%		59.76%		59.93%	
<i>n</i>	354		351		348		297	
Cluster	19		19		19		19	

Note. t0 = beginning of first grade. ⁺*p* < .10. * *p* < .05. ** *p* < .01. *** *p* < .001.

order to test for the significance of the indirect effects. The results (models not displayed) confirmed the insignificance of the indirect effects and therefore indicate direct-only nonmediation for all three indicators of teacher feedback examined in this study.

6. Discussion

Contributing to the body of research on the role of teachers' expectations in students' academic achievements, this study focused on the mediating mechanisms and empirically traced all three steps of the

assumed self-fulfilling prophecy process. Our first goal was to clarify whether associations between teacher expectations and later student achievement occurred due to differences in teacher behavior. The second aim was to resolve open questions surrounding the role of teacher feedback as a mediator of teacher expectancy effects. In this regard, we examined the associations between teacher expectations, teacher feedback practice and student learning.

In summary, the results showed that teacher expectations were inaccurate to some extent. Further, this inaccuracy in teacher expectations significantly predicted students' end-of-year achievement, even when prior achievement, general cognitive abilities, motivation, and student background characteristics were controlled. Specifically, inaccurately high expectations were related to greater achievement gains in reading and mathematics, whereas inaccurately low expectations were related to lower achievement gains in reading only. In addition, teacher feedback varied with teacher expectations but did not substantially mediate teacher expectancy effects.

The results are in line with existing findings indicating that teacher expectations are partly inaccurate (Research Question 1). That is, in accordance with Hypothesis 1, teacher expectations differ from actual student achievement, general cognitive abilities, and motivation. The shared variance between student characteristics and teacher expectations in the German language domain and in mathematics amounted to approximately 35%, corresponding to estimates reported in the available meta-analyses (Hoge & Coladarci, 1989; Südkamp et al., 2012). When using dichotomous inaccuracy categories, we defined teacher expectations to be accurate in a range of one standard deviation below and above the mean. This approach resulted in 65–70% of teacher expectations being defined as accurate. However, the exact proportion depended on the tolerance range chosen. With a tolerance range of 0.5 standard deviations, for example, only 40% of the teacher expectations were categorized as accurate. The exact proportions of the categories should thus be interpreted with caution. As the robustness checks of all analyses with different tolerance ranges of teacher expectation accuracy revealed, the results and the conclusions did not differ substantially depending on the definition of this criterion.

Teacher expectation inaccuracy was also significantly associated with two dimensions of teacher feedback (Research Question 2). In line with Hypotheses 2a and 2b, compared to similar-achieving lower-expectancy classmates, higher-expectancy students received more performance feedback than behavioral feedback and somewhat more positive performance feedback than negative performance feedback. This finding generally supports the assumption that teachers communicate their expectations through different feedback practices. The frequency of further elaboration provided by teachers in the case of negative performance feedback, however, did not increase with inaccurately higher expectations, which contradicted Hypothesis 2c.

The results further indicate that teacher expectations can result in a self-fulfilling prophecy (Research Question 3). Associations between initial teacher expectation inaccuracy and later student achievements were found for the German language and mathematical domains, but the regression coefficient was larger in the language domain ($\beta = 0.21$ compared to $\beta = 0.09$). Correspondingly, whereas we found positive and negative teacher expectancy effects in the language domain, we observed only positive expectancy effects in mathematics. Therefore, the results support Hypothesis 3a for both domains but support Hypothesis 3b for the German language domain only. There could be various reasons for the differences between domains. One possible explanation might be that the linguistic domain provides more room for interpretation in the evaluation of achievement. The scope for interpretation is smaller in the case of mathematics, in which most tasks have objectively correct and incorrect responses. Another reason might be related to the different levels of preknowledge that students bring with them when entering first grade. Whereas many children can successfully count to ten or twenty and have mastered the first rules of arithmetic (cf. Deutscher & Selter, 2013), only a few already can read

when they enter school (cf. Juska-Bacher, 2013). As a result, the growth in reading achievement is much steeper than that in mathematics, and teachers (and their expectations) might therefore have a stronger impact on reading development.

The sizes of teacher expectancy effects ($\beta = 0.21$ in the German language domain and $\beta = 0.09$ in mathematics) correspond closely to the findings of earlier studies (effect sizes of 0.10–0.20; see, e.g., Jussim et al., 2009). According to the effect size guidelines for research on individual differences based on 708 meta-analyses, the effects were small to medium (Gignac & Szodorai, 2016). However, it seems necessary to consider the short time period of one school year covered in the empirical study. Teacher expectancy effects might accumulate and become stronger over longer periods (Rubie-Davies et al., 2014). Furthermore, since teacher expectations have been found to be systematically biased toward different groups of students (based on student socioeconomic background, country of origin, gender, or disability status; e.g., Lorenz et al., 2016; Shifrer, 2016), the observed expectancy effects might contribute to educational inequalities.

The mediation analyses yielded no signs of a strong mediation of teacher expectancy effects by teacher feedback (Research Question 4) and thus contradicted Hypothesis 4. Following the classification of mediation and nonmediation by Zhao et al. (2010), the results indicate direct-only nonmediation. Apart from a positive association between positive performance feedback and mathematical achievement gains, no significant correlations between teacher behavior and student learning were found. Nevertheless, the results support the association of teacher expectations with teacher feedback. One reason why the analyzed characteristics of teacher feedback did not predict student achievement gains could be the difficulty of adequately measuring high-quality feedback. Another reason for the weak support for a mediation of the effects of teacher expectations by teacher feedback might be that feedback is one of numerous channels through which teachers communicate their expectations.

6.1. Limitations and future research

Although the design of the current study allowed us to address several shortcomings of earlier research, the study also has limitations that should be addressed in future research. One limitation concerns the sample. Since teachers and students participated in the study voluntarily, the sample might be selective in terms of teacher engagement and students' socioeconomic background. A comparison of the average HISEI values in the sample ($M = 52.51$) to the population average in North Rhine-Westphalia ($M = 48.10$; Richter, Kuhl, & Pant, 2012) revealed a positive bias in the student sample. Regarding teachers, one would expect highly engaged teachers to participate more often in such a study and to be particularly eager to evaluate students accurately. As a result, the extent of inaccuracy in teacher expectations and its association with student learning might be somewhat underestimated in our study.

In addition, as the current study examined data from teachers and students in German primary schools, the conclusions are restricted to this context, more precisely, to first grade classrooms in North Rhine-Westphalia. Replication studies based on data from other countries, different grade levels (e.g., after a school transition) or specific groups of students (e.g., with different levels of self-perceptions of their ability) are important next steps to better understand the processes underlying teacher expectancy effects and ensure external validity of the current findings.

Furthermore, the subsample including information on teacher behavior was rather small. Only 19 teachers (with 354 students) agreed to be videotaped during lessons. In addition, based on the intraindividual measures of teacher feedback, we had to exclude some students from the analyses because they did not receive any feedback, performance feedback or negative performance feedback during the videotaped lessons. Therefore, the statistical power of the analyses was limited, and

we could use only a restricted number of explanatory variables in the models because of the limited degrees of freedom. Additionally, we could observe only extracts of teacher-student interactions. This limitation applies to the types of interactions, as well as to their duration. Specifically, we could examine only three characteristics of teacher feedback in four school lessons per class and were not able to consider the quality and appropriateness of students' preceding answers and behaviors. The main reason for this limitation is that video ratings are complex and expensive. However, in light of these restrictions, it is all the more remarkable that the analyses revealed statistically significant associations among teacher expectations, teacher behavior and, partly, student learning. Nevertheless, further studies should focus on the mediating mechanisms with larger sample sizes and additional dimensions of teacher behavior, as well as finer-grained measures of high-quality teacher feedback. Additional dimensions of teacher behavior may include aspects covered by the four-factor model (Rosenthal, 1973), teacher behaviors that are discussed in the context of enhancing student motivation (e.g., TARGET framework; Jussim et al., 2009) as well as other channels for the mediation of teacher expectations such as non-verbal communication (Babad, 2009). Finer-grained ratings of high-quality feedback should, for example, provide more detailed information about the three aspects of effective feedback, that is, whether the feedback contains information about (1) the learning goal, (2) how well a student is doing, and (3) the next steps to be undertaken to reach the learning goal (Hattie & Timperley, 2007). In addition, ratings should include information about the quality of students' preceding answers and behaviors.

A further limitation concerns the residual approach that was applied to identify teacher expectation inaccuracy. The unexplained variance in teacher expectations might not necessarily reflect inaccuracy but could at least partly be the result of measurement error or differences in unobserved student characteristics. First, with regard to measurement error, the risk of overestimating inaccuracy cannot be fully ruled out in field studies because not even the best instruments to assess learning-related student characteristics will represent students' true values of achievement, general cognitive abilities or motivation without error. However, as long as the potential misestimations varied randomly around the true values, the problem of measurement error should not have biased the examined relationships between teacher expectations, teacher feedback and student achievement systematically. Second, the unexplained variance in teacher expectations could partly be the result of unobserved student characteristics. If, for example, teachers had observed differences in student achievement and skills not captured by the various instruments used in this study, the residuals would not represent teacher expectation inaccuracy; instead, they could reflect accurate influences of unobserved student characteristics (cf. Lorenz, 2018). However, the various learning-related student characteristics (including general cognitive abilities and motivation) and the way in which the data were collected in our study should have minimized this problem. Since teacher-student contact prior to the first survey wave was minimal, it is fair to assume not only that teacher expectations had not yet affected initial student achievement but also that teacher expectations had not yet been determined by comprehensive evaluations of unmeasured individual student characteristics. This assumption is supported by a study that showed that ethnic bias in teacher expectations varied with teachers' stereotypes (Lorenz, 2019). Thus, we can assume that a substantial portion of the residual scores for teacher expectations emerged due to teachers' application of schemata and not due to the processing of individual student information.

6.2. Conclusion

Our study provides evidence for the relevance of teacher expectancy

effects for students' learning during the first grade in two different school domains and, in part, for the communication of teacher expectations through teachers' feedback practices. Although the examined characteristics of teacher feedback did not mediate teacher expectancy effects, the effectiveness of high-quality feedback in supporting student learning remains important (e.g., Hattie, 2009). High-quality feedback is closely related to learning goals and informs students about their progress as well as about how to proceed to reach their goals (Hattie & Timperley, 2007). To be effective, such feedback requires that teachers and students regularly set appropriate goals together. Such goals should be specific, clear and challenging but achievable; they should be reviewed and updated regularly with each student (Clarke, Timperley, & Hattie, 2003). Based on our results, teacher trainings that focus on formative high-quality feedback should enable teachers to provide supporting feedback to all their students, independent of what achievement they perceive or expect from these students.

Furthermore, our findings support the idea that students learn best if teacher demands are slightly higher than students' actual skills. This finding implies that high expectations might be beneficial for all students. Therefore, it seems worthwhile to inform teachers not only about expectancy effects in general but also about the positive effects of high expectations in particular. In addition, teachers may be encouraged to form high expectations for all their students. This is particularly important, as teacher expectancy effects reflect mostly unconscious processes. Hence, the prevention of such effects might not be entirely realistic. A central challenge in the implementation of high expectations for all students, however, may be to reduce biasing influences associated with students' gender, family backgrounds and disabilities, as current research indicates group-specific biases in teacher expectations (e.g., Hurwitz et al., 2007; Lorenz et al., 2016). This approach involves knowledge about stereotypes and strategies to reduce their influence when forming expectations.

An intervention study by Rubie-Davies, Peterson, Sibley, and Rosenthal (2015) showed that a focus on high overall expectations for all students might be a promising route. In four workshops spread over two months, the authors trained the participating teachers in the practices of teachers who had high expectations for all their students (e.g., providing goal-related feedback; Rubie-Davies et al., 2007). These workshops were supplemented by periods in which the teachers implemented the learned practices in their classrooms, self-analyses of their video-taped classroom practices and follow-up meetings with the researchers as project partners. After the intervention program, the students who were taught by teachers from the intervention group had gained more competencies in mathematics over the year than the students from the control group. This and similar training programs are valuable and promising attempts to benefit from the positive effects of high expectations and may support teachers in providing high-quality feedback to all of their students.

Funding

The authors acknowledge financial support from the Federal Ministry of Education and Research of Germany (project number 01JC1117). The responsibility for the content of this publication lies with the authors.

Author note

We would like to thank Petra Stanat for her valuable comments on an earlier version of this manuscript. Her remarks helped very much to further improve the text.

Appendix A. Descriptive statistics of all variables in the main sample and the subsample

Variable	Main sample (N = 1026)				Subsample (n = 354)			
	M	SD	Min	Max	M	SD	Min	Max
Student achievement t0								
Phonological awareness	13.25	4.64	1	26	13.40	4.45	2	26
Reading ^a	23.22	26.43	0	156	20.15	23.46	0	147
Mathematics	27.00	8.06	0	53	26.62	8.43	0	53
Student achievement t2								
Phonological awareness	20.82	4.61	3	26	20.77	4.55	7	26
Reading	107.47	37.65	3	158	109.16	35.81	7	158
Mathematics	38.61	7.03	17	53	38.71	6.94	19	53
Student general cognitive abilities								
Deductive reasoning	6.61	2.90	0	12	6.50	2.96	0	12
Working memory	3.11	1.70	0	8	2.88	1.66	0	7
Student motivation								
Enjoyment of learning	1.53	0.38	0	2	1.54	0.37	0	2
Effort	1.69	0.28	0.31	2	1.69	0.28	0.62	2
Student background								
Age	6.45	0.33	5.58	8.08	6.48	0.35	5.67	8
HISEI	52.51	19.44	14.21	88.96	52.03	19.69	14.21	88.96
	%				%			
Gender: Female	48.11				47.18			
Parental education: Abitur or higher	49.41				49.86			
Immigrant status	35.55				38.40			

Note. Descriptive statistics are based on the imputed data (pooled according to Rubin's rules; Rubin, 1987). For descriptive statistics on teacher expectations, teacher expectation inaccuracy and teacher feedback, see Table 2 t0 = beginning of first grade. t2 = end of first grade.

^a Because there are usually some students who have already learned how to read before they entered school, whereas most students have not, the reading variable naturally shows a very large standard deviation at the beginning of first grade. Robustness checks excluding this variable from the regression models, displayed in Tables 4–6, confirmed that the results are reliable.

Appendix B. Excerpt from the teacher questionnaire measuring teacher expectations (cf. Gentrup, Rjosk, Stanat, & Lorenz, 2018)

Please evaluate how the child is likely to develop with regard to the following skills during the upcoming school year by comparing the child to the class average.						
		The child will perform in this area ...				
		much worse than...	a little worse than ...	just as well as ...	slightly better than ...	much better than...
		... the class average.				
a)	linguistic skills (e.g., vocabulary and sentence structure)	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
b)	literacy skills (e.g., reading and writing words)	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
c)	mathematical skills (e.g., dealing with numbers and quantities)	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅

Compared to his/her fellow students, how well do you expect this child to perform at the end of the school year?						
		far below the class average	below the class average	near the class average	above the class average	far above the class average
a)	in mathematics	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
b)	in German language	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅

Appendix C. Statistics on the agreement between rater coding and master coding of videotaped teacher-student interactions

Variable	Video-rater ID					
	1	2	3	4	5	6
First training period						
Content-related agreement (<i>kappa</i>) without predefined subsequences ^a	.66	–	.46	.70	–	.48
Number of subsequently coded school lessons	6	0	4	1	0	3
Second training period						
Content-related agreement (<i>kappa</i>) without predefined subsequences ^a	.69	.70	–	–	.62	.63
Content-related agreement (<i>kappa</i>) with predefined subsequences ^b	.96	.97	–	–	.97	.96
Bivariate correlation between rater coding and master coding regarding ...						
... performance relatedness of FB ^c	.86	.97	–	–	.93	.95
... valence of performance FB	.78	.92	–	–	.77	.81
... elaborateness of negative performance FB	.70	.77	–	–	.78	.87
Number of subsequently coded school lessons	17	12	0	0	12	10

Note. Raters 2 and 5 did not participate in the first training period, whereas raters 3 and 4 were no longer available for the second training period.

^a To calculate Cohen's kappa, every interaction sequence was split into 100-msec sections. The resulting estimate indicates the agreement between a rater coding and the master coding, considering differences in the timing of subsequences.

^b This measure is based on content-related ratings of 168 predefined subsequences and indicates the agreement beyond issues of the timing of subsequences.

^c FB = feedback.

References

- Archambault, I., Janosz, M., & Chouinard, R. (2012). Teacher beliefs as predictors of adolescents' cognitive engagement and achievement in mathematics. *The Journal of Educational Research*, 105(5), 319–328. <https://doi.org/10.1080/00220671.2011.629694>.
- Artelt, C., Blossfeld, H.-P., Faust, G., Roßbach, H.-G., & Weinert, S. (2013). *Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter (BiKS-3-10)* [Educational processes, competence development and selection decisions in preschool and school age (BiKS-3-10)]. IQB – Institut zur Qualitätsentwicklung im Bildungswesen https://doi.org/10.5159/IQB_BIKS_3_10_v4 Version: 4.
- Babad, E. (2009). *The social psychology of the classroom*. New York: Taylor & Francis.
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachandran (Vol. Ed.), *Encyclopedia of human behavior: Vol. 4*, (pp. 71–81). New York: Academic Press.
- Bäuerlein, K., Beinicke, A., Berger, N., Faust, G., Jost, M., & Schneider, W. (2012). *FIPS fähigkeitsindikatoren Primarschule: Ein computerbasiertes Diagnoseinstrument zur Erfassung der Lernausgangslage und der Lernentwicklung von schulanfängern. Manual* [Performance indicators in primary schools: A computer-based diagnostic tool for detecting the initial learning capabilities and the learning development of children starting school]. (Göttingen: Hogrefe).
- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 25–48). New York: Springer Science + Business Media.
- Baumert, J., & Maaz, K. (2006). Das theoretische und methodische Konzept von PISA zur Erfassung sozialer und kultureller Ressourcen der Herkunftsfamilie: Internationale und nationale Rahmenkonzeption [The theoretical and methodological concept of PISA to assess social and cultural resources of the family of origin: International and national conceptual framework]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, D. (2013). The impact of teachers' expectations on students' educational opportunities in the life course: An empirical test of a subjective expected utility explanation. *Rationality and Society*, 25(4), 422–469. <https://doi.org/10.1177/1043463113504448>.
- de Boer, H., Bosker, R. J., & van der Werf, M. P. C. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, 102(1), 168–179. <https://doi.org/10.1037/a0017289>.
- de Boer, H., & van der Werf, M. P. C. (2015). Influence of misaligned parents' aspirations on long-term student academic performance. *Educational Research and Evaluation*, 21(3), 232–257. <https://doi.org/10.1080/13803611.2015.1039548>.
- Bohn, C. M., Roehrig, A. D., & Pressley, M. (2004). The first days of school in the classrooms of two more effective and four less effective primary-grades teachers. *The Elementary School Journal*, 104(4), 269–287. <https://doi.org/10.1086/499753>.
- van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049–1064. <https://doi.org/10.1080/10629360600810434>.
- Chen, Y.-H., Thompson, M. S., Kromrey, J. D., & Chang, G. H. (2011). Relations of student perceptions of teacher oral feedback with teacher expectancies and student self-concept. *The Journal of Experimental Education*, 79(4), 452–477. <https://doi.org/10.1080/00220973.2010.547888>.
- Chow, S. L. (1987). Some reflections on Harris and Rosenthal's 31 meta-analyses. *Journal of Psychology*, 121(1), 95–100. <https://doi.org/10.1080/00223980.1987.9712648>.
- Clarke, S., Timperley, H. S., & Hattie, J. (2003). *Unlocking formative assessment: Practical strategies for enhancing students' learning in the primary and intermediate classroom*. Auckland, New Zealand: Hodder Moa Beckett.
- Dean, C. B., Hubbell, E. R., Pitler, H., & Stone, B. (2012). *Classroom instruction that works: Research-based strategies for increasing student achievement* (2nd ed.). Denver: McREL.
- Deutscher, T., & Selzer, C. (2013). Frühe mathematische Bildung – forschungsbefunde und Förderkonzepte [Early mathematical education – Empirical evidence and training approaches]. In M. Stamm, & D. Edelmann (Eds.), *Handbuch frühkindliche bildungs-forschung* (pp. 543–556). Wiesbaden: Springer VS.
- Dietrich, J., Dicke, A.-L., Kracke, B., & Noack, P. (2015). Teacher support and its influence on students' intrinsic value and effort: Dimensional comparison effects across subjects. *Learning and Instruction*, 39, 45–54. <https://doi.org/10.1016/j.learninstruc.2015.05.007>.
- Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology*, 33, 1–38.
- Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 155–171). New York: Guilford Press.
- Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K., & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology*, 41, 1–12. <https://doi.org/10.1016/j.cedpsych.2014.10.006>.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102(4), 652–670.
- Ganzeboom, H. B. G. (2010). (May). *A new international socio-economic index (ISEI) of occupational status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007*. Lisbon, Portugal: Annual Conference of the International Social Survey Programme Paper presented at the.
- Ganzeboom, H. B. G., Graaf, P. M. D., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1–56.
- Gentrup, S., & Rjosk, C. (2018). Pygmalion and the gender gap: Do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling? *Educational Research and Evaluation*, 24(3–5), 295–323. <https://doi.org/10.1080/13803611.2018.1550840>.
- Gentrup, S., Rjosk, C., Stanat, P., & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen [Teachers' perceptions of students' motivation and learning behaviour and their role in biased teacher achievement expectations]. *Zeitschrift für Erziehungswissenschaft*, 21, 867–891. <https://doi.org/10.1007/s11618-018-0806-2>.
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>.
- Gill, S., & Reynolds, A. J. (1999). Educational expectations and school achievement of urban African American children. *Journal of School Psychology*, 37(4), 403–424. [https://doi.org/10.1016/S0022-4405\(99\)00027-8](https://doi.org/10.1016/S0022-4405(99)00027-8).
- Harris, M., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97(3), 363–386 0033-2909/85/S00.75.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101(3), 662–670. <https://doi.org/10.1037/a0014306>.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement:

- A review of literature. *Review of Educational Research*, 59(3), 297–313. <https://doi.org/10.3102/00346543059003297>.
- Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social Psychology of Education*, 20, 471–490. <https://doi.org/10.1007/s11218-017-9384-z>.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam, & J. Neyman (Vol. Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability: Vol. 1*, (pp. 221–233). Berkeley: University of California Press.
- Hurwitz, J. T., Elliott, S. N., & Braden, J. P. (2007). The influence of test familiarity and student disability status upon teachers' judgments of students' test performance. *School Psychology Quarterly*, 22(2), 115–144.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509–527.
- Jenkins, L. N., & Demaray, M. K. (2016). Teachers' judgments of the academic achievement of children with and without characteristics of inattention, impulsivity, and hyperactivity. *Contemporary School Psychology*, 20(2), 183–191. <https://doi.org/10.1007/s40688-015-0073-7>.
- Johnson, T. J., Feigenbaum, R., & Weiby, M. (1964). Some determinants and consequences of the teacher's perception of causation. *Journal of Educational Psychology*, 55(5), 237–246. <https://doi.org/10.1037/h0043389>.
- Jose, P. E. (2013). *Doing statistical mediation and moderation*. London: Guilford Press.
- Jungbluth, P. (1993). Pygmalion and the effectiveness of 'Black' schools: Teachers' stereotypes and hidden goal differentiations towards ethnic minorities. *Tijdschrift Voor Onderwijsresearch*, 18(2), 99–110.
- Jungbluth, P. (1994). Lehrererwartungen und Ethnizität. Innerschulische Chancendeterminanten bei Migrantenschülern in den Niederlanden [Teacher expectations and ethnicity. The educational opportunities of adolescent migrants in The Netherlands]. *Zeitschrift für Pädagogik*, 40(1), 113–125 doi:urn:nbn:de:0111-pedocs-108314.
- Juska-Bacher, B. (2013). Leserelevante Kompetenzen und ihre frühe Förderung [Reading-related competencies and their early training]. In M. Stamm, & D. Edelmann (Eds.), *Handbuch frühkindliche bildungsforschung* (pp. 485–500). Wiesbaden: Springer VS.
- Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review*, 93(4), 429–445. <https://doi.org/10.1037/0033-295X.93.4.429>.
- Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Vol. Ed.), *Advances in experimental social psychology: Vol.28*, (pp. 281–388). San Diego: Academic Press.
- Jussim, L., & Harber, K. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155. <https://doi.org/10.1207/s15327957pspr0902.3>.
- Jussim, L., Robustelli, S., & Cain, T. (2009). Teacher expectations and self-fulfilling prophecies. In K. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 349–379). New York: Routledge.
- Kristen, C., Stanat, P., Kogan, I., Lorenz, G., Gentrup, S., & Rahmann, S. (2018a). *Kompetenzerwerb und lernvoraussetzungen (KuL) [competence acquisition and learning preconditions (KuL)]*. IQB – Institut zur Qualitätsentwicklung im Bildungswesen <https://doi.org/10.5159/IQB.KuL.v1> Version: 1.
- Kristen, C., Stanat, P., Kogan, I., Lorenz, G., Gentrup, S., & Rahmann, S. (2018b). *Kompetenzerwerb und lernvoraussetzungen (KuL) [competence acquisition and learning preconditions (KuL)]*. Forschungsdatenzentrum Bildung am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. Observational data <https://doi.org/10.7477/287:1.0>.
- Kuklinski, M. R., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development*, 72, 1554–1578.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>.
- Lorenz, G. (2018). *Selbsterfüllende Prophezeiungen in der Schule. Leistungserwartungen von Lehrkräften und Kompetenzen von Kindern mit Zuwanderungshintergrund [Self-fulfilling prophecies in school. Teacher expectations and competences of immigrant children]*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lorenz, G. (2019). *Subtle discrimination? Stereotypes among teachers trigger ethnic bias in their achievement expectations but are unrelated to ethnic achievement gaps.* (in preparation).
- Lorenz, G., Gentrup, S., Kristen, C., Stanat, P., & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine untersuchung systematisch verzerrter lehrererwartungen [stereotypes among teachers? A study of systematic bias in teacher expectations]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(1), 89–111. <https://doi.org/10.1007/s11577-015-0352-3>.
- Ma, X. (2001). Participation in advanced mathematics: Do expectation and influence of students, peers, teachers, and parents matter? *Contemporary Educational Psychology*, 26(1), 132–146. <https://doi.org/10.1006/ceps.2000.1050>.
- Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, 72, 791–809. <https://doi.org/10.1037/0022-3514.72.4.791>.
- van Matre, J. C., Valentine, J. C., & Cooper, H. M. (2000). Effect of students' after-school activities on teachers' academic expectancies. *Contemporary Educational Psychology*, 25, 167–183. <https://doi.org/10.1006/ceps.1998.0999>.
- McKown, C., & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology*, 32(1), 159–184. <https://doi.org/10.1111/j.1559-1816.2002.tb01425.x>.
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, 46(3), 235–261. <https://doi.org/10.1016/j.jsp.2007.05.001>.
- Mehmetoglu, M. (2018). medsem: a Stata package for statistical mediation analysis. *International Journal of Computational Economics and Econometrics*, 8(1), 63–78.
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. <https://doi.org/10.1016/j.tate.2017.02.021>.
- Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review*, 8(2), 193–210. <https://doi.org/10.2307/4609267>.
- Morris, E. W. (2005). From “middle class” to “trailer trash:” Teachers' perceptions of white students in a predominately minority school. *Sociology of Education*, 78(2), 99–121. <https://doi.org/10.1177/003804070507800201>.
- Muntoni, F., & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—the role of teachers' gender stereotypes. *Contemporary Educational Psychology*, 54, 212–220. <https://doi.org/10.1016/j.cedpsych.2018.06.012>.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- OECD (2003). *Literacy skills for the world of tomorrow: Further results from PISA 2000*. Paris: OECD Publishing.
- Pellegrini, R. J., & Hicks, R. A. (1972). Prophecy effects and tutorial instruction for the disadvantaged child. *American Educational Research Journal*, 3, 413–419. <https://doi.org/10.2307/1161757>.
- Pendry, L. (2015). Social cognition. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.), *An introduction to social psychology* (pp. 93–122). (6 ed.). Chichester: Wiley.
- Peterson, E. R., Rubie-Davies, C. M., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, 42, 123–140.
- Rauer, W., & Schuck, K. D. (2004). *Feess 1-2: Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen [FEES 1-2: Questionnaire for emotional and social school experiences of first- and second-grade students]*. Göttingen: Beltz Test GmbH.
- Ready, D. D., & Chu, E. M. (2015). Sociodemographic inequality in early literacy development: The role of teacher perceptual accuracy. *Early Education & Development*, 26(7), 970–987. <https://doi.org/10.1080/10409289.2015.1004516>.
- Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, 48(2), 335–360. <https://doi.org/10.3102/0002831210374874>.
- Richter, D., Kuhl, P., & Pant, H. A. (2012). Soziale disparitäten [social disparities]. In P. Stanat, H. A. Pant, K. Böhme, & D. Richter (Eds.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (pp. 191–207). Münster: Waxmann.
- Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society*, 26(2), 290–322. <https://doi.org/10.1177/0891243211434614>.
- Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, 81(4), 493–529. <https://doi.org/10.3102/0034654311421793>.
- Rosenthal, R. (1973). On the mediation of pygmalion effects: A four factor “theory” *Papua New Guinea Journal of Education*, 9, 1–12.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 16–20.
- Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology*, 77, 289–306. <https://doi.org/10.1348/000709906X101601>.
- Rubie-Davies, C. M., Hattie, J. A. C., Townsend, M. A. R., & Hamilton, R. J. (2007). Aiming high: Teachers and their students. In V. N. Galwey (Ed.), *Progress in educational psychology research* (pp. 65–91). Hauppauge, NY: Nova Publisher.
- Rubie-Davies, C. M., & Peterson, E. R. (2011). Teacher expectations and beliefs. Influences on the socioemotional environment of the classroom. In C. M. Rubie-Davies (Ed.), *Educational psychology. Concepts, research and challenges* (pp. 135–149). New York: Routledge.
- Rubie-Davies, C. M., & Peterson, E. R. (2016). Relations between teachers' achievement, over- and underestimation, and students' beliefs for Māori and Pākehā students. *Contemporary Educational Psychology*, 47, 72–83. <https://doi.org/10.1016/j.cedpsych.2016.01.001>.
- Rubie-Davies, C. M., Peterson, E. R., Sibley, C. G., & Rosenthal, R. (2015). A teacher expectation intervention: Modelling the practices of high expectation teachers. *Contemporary Educational Psychology*, 40, 72–85. <https://doi.org/10.1016/j.cedpsych.2014.03.003>.
- Rubie-Davies, C. M., Weinstein, R. S., Huang, F. L., Gregory, A., Cowan, P. A., & Cowan, C. P. (2014). Successive teacher expectation effects across the early school years. *Journal of Applied Developmental Psychology*, 35(3), 181–191. <https://doi.org/10.1016/j.appdev.2014.03.006>.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.
- Schenke, K., Nguyen, T., Watts, T. W., Sarama, J., & Clements, D. H. (2017). Differential effects of the classroom on african american and non-african american's mathematics achievement. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000165>.
- Schuchardt, C., & Dunkake, I. (2014). Schichtspezifische Stereotype unter angehenden Lehrkräften [Social class stereotypes among teacher trainees]. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 34(1), 89–107. <https://doi.org/10.3262/ZSEI1401089>.
- Shiffrer, D. (2013). Stigma of a label: Educational expectations for high school students labeled with learning disabilities. *Journal of Health and Social Behavior*, 54(4),

- 462–480. <https://doi.org/10.1177/0022146513503346>.
- Shifrer, D. (2016). Stigma and stratification limiting the math course progression of adolescents labeled with a learning disability. *Learning and Instruction*, 42, 47–57. <https://doi.org/10.1016/j.learninstruc.2015.12.001>.
- Snow, R. E. (1969). Review of "pygmalion in the classroom" by rosenthal and jacobson. *Current Directions in Psychological Science*, 14, 197–199.
- Spinath, B., & Steinmayr, R. (2008). Longitudinal analysis of intrinsic motivation and competence beliefs: Is there a relation over time? *Child Development*, 79(5), 1555–1569.
- Stahl, N. (2007). Schülerwahrnehmung und -beurteilung durch Lehrkräfte [Student perception and assessment by teachers]. In H. Ditton (Ed.). *Kompetenzaufbau und Laufbahnen im Schulsystem* (pp. 171–198). Münster: Waxmann.
- StataCorp, L. L. C. (1985–2017). *Stata/SE 15.1 [computer software]*. Texas: College Station.
- StataCorp (2017). *Stata user's guide. Release, Vol. 15*. Texas: Stata Press.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>.
- Thorndike, R. L. (1968). Review of robert rosenthal and lenore Jacobson, pygmalion in the classroom. *American Educational Research Journal*, 5(4), 708–711. <https://doi.org/10.2307/1162010>.
- Timmermans, A. C., de Boer, H., & van der Werf, M. P. C. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education*, 19(2), 217–240. <https://doi.org/10.1007/s11218-015-9326-6>.
- Timmermans, A. C., Kuyper, H., & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, 85(4), 459–478. <https://doi.org/10.1111/bjep.12087>.
- Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education*, 45, 73–82. <https://doi.org/10.1016/j.tate.2014.09.006>.
- Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, 25, 1051–1060. <https://doi.org/10.1016/j.tate.2009.04.002>.
- Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, 24(3–5), 124–179. <https://doi.org/10.1080/13803611.2018.1548798>.
- Weinstein, R. S. (1983). Student perceptions of schooling. *The Elementary School Journal*, 83(4), 286–312.
- Weiß, R. H., & Osterland, J. (1997). *Grundintelligenztest skala 1: CFT 1. Handanweisung [culture fair intelligence test 1: CFT1. Manual] (5., revidierte auflage ed.)*. Göttingen: Hogrefe.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1–25.
- Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering baron and kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37, 197–206. <https://doi.org/10.1086/651257>.
- Zhou, J., & Urhahne, D. (2013). Teacher judgment, student motivation, and the mediating effect of attributions. *European Journal of Psychology of Education*, 28(2), 275–295.